



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) **EP 1 266 906 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication:
18.12.2002 Bulletin 2002/51

(51) Int Cl.7: **C07K 14/47**, C07K 16/18,
C12N 15/12, C12N 5/10,
C12Q 1/68, G01N 33/68

(21) Application number: **02254112.2**

(22) Date of filing: **13.06.2002**

(84) Designated Contracting States:
**AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
MC NL PT SE TR**
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **13.06.2001 US 298012 P**

(71) Applicant: **Millennium Pharmaceuticals, Inc.
Cambridge, Massachusetts 02139 (US)**

(72) Inventor: **Curtis, Rory A.J.
Framingham Massachusetts 01702 (US)**

(74) Representative: **Lunt, Mark George Francis et al
Harrison Goddard Foote,
Fountain Precinct,
Leopold Street
Sheffield S1 2QD (GB)**

(54) **46584, a human transporter family member and uses therefor**

(57) The invention provides isolated nucleic acids molecules, designated 46584 nucleic acid molecules, which encode transporter family members. The invention also provides antisense nucleic acid molecules, recombinant expression vectors containing 46584 nucleic acid molecules, host cells into which the expression vectors have been introduced, and nonhuman transgenic

animals in which a 46584 gene has been introduced or disrupted. The invention still further provides isolated 46584 proteins, fusion proteins, antigenic peptides and anti-46584 antibodies. Diagnostic and therapeutic methods utilizing compositions of the invention are also provided.

EP 1 266 906 A1

Description**Background of the Invention**

[0001] Cellular membranes differentiate the contents of a cell from the surrounding environment. Membranes also serve as effective barriers against the unregulated influx of hazardous or unwanted compounds, and the unregulated efflux of desirable compounds. However, the cell does need a supply of desired compounds and removal of waste products. Transport proteins which are embedded (singly or in complexes) in the cellular membrane (reviewed by Oh and Amidon (1999) in *Membrane Transporters as Drug Targets*, ed. Amidon and Sadée, Kluwer Academic/Plenum Publishers, New York, Chapter 1) are major providers of these functions. There are two general classes of membrane transport proteins: channels or pores, and transporters (also known as carriers or permeases). Channels and transporters differ in their translocation mechanisms. Channels are hydrophilic group-lined protein tunnels whose opening by a regulatory event allow free, rapid passage of their charge-, size-, and geometry-selected small ions down their concentration gradients. Transporters specifically and selectively bind the molecules they move, some with and some against their concentration gradients, across membranes. The binding mechanism causes the action of transporters to be slow and saturable.

[0002] Transport molecules are specific for a particular target solute or class of solutes, and are also present in one or more specific membranes. Transport molecules localized to the plasma membrane permit an exchange of solutes with the surrounding environment, while transport molecules localized to intracellular membranes (e.g., membranes of the mitochondrion, peroxisome, lysosome, endoplasmic reticulum, nucleus, or vacuole) permit import and export of molecules from organelle to organelle or to the cytoplasm. For example, in the case of the mitochondrion, transporters in the inner and outer mitochondrial membranes permit the import of sugar molecules, calcium ions, and water (among other molecules) into the organelle and the export of newly synthesized ATP to the cytosol.

[0003] Transporters can move molecules by two types of processes. In one process, "facilitated diffusion," transporters move molecules with their concentration gradients. In the other process, "active transport," transporters move molecules against their concentration gradients. Active transport to move a molecule against its gradient requires energy, in contrast to facilitated diffusion, which does not require energy.

[0004] Transporters play important roles in the ability of the cell to regulate homeostasis, to grow and divide, and to communicate with other cells, e.g., to transport signaling molecules, such as hormones, reactive oxygen species, ions, neurotransmitters or vitamins. A wide variety of human diseases and disorders are associated with defects in transporter or other membrane transport molecules, including certain types of liver disorders (e.g., due to defects in transport of long-chain fatty acids (Al Odaib *et al.* (1998) *New Eng. J. Med.* 339:1752-1757), hyperlysinemia (mitochondrial lysine transport defect (Oyanagi *et al.* (1986) *Inherit. Metab. Dis.* 9:313-316)), and cataract (Wintour (1997) *Clin. Exp. Pharmacol. Physiol.* 24(1):1-9).

[0005] There are over 30 families of secondary transporters, also known as solute carriers or SLC (reviewed by Berger, *et al.* (2000) in *The Kidney: Physiology and Pathophysiology*, eds. Seldin DW and Giebisch G., Lippincott, Williams & Wilkins, Philadelphia 1:107-138; also see www.gene.ucl.ac.uk/nomenclature for human proteins). The SLC families are classified on the basis of the type of molecules they transport.

[0006] The SLC2 family members are facilitated-diffusion sugar transporters. They transport various monosaccharides, e.g., glucose, fructose, galactose, and myo-inositol across the cell membrane, and thus participate in glucose homeostasis. The cDNAs of at least five distinct glucose transporters (Glut1 to Glut4 and Glut7) and one fructose transporter (Glut5) have been isolated. These transporters differ in tissue expression, kinetics, and substrate specificity, but share a similar twelve-membrane-spanning domain structure.

[0007] The SLC21 and 22 families transport organic ions. Some members of the SLC21 family transport organic anions, others transport prostaglandins. Some members of the SLC22 family transport organic anions, others transport organic cations, and some may transport either type of ion. Members of these families are polyspecific, but all contain twelve transmembrane domains, often consisting of α -helices. They participate in activities as diverse as intestinal or hepatic absorption of metabolites, renal reabsorption of cations or excretion of cations. Members of these families also transport a wide variety of drugs and xenobiotics, many of which are harmful to the body. In addition, organic ion transporters are responsible for the transport of the metabolites of most lipophilic compounds, e.g., sulfate and glucuronide conjugates (Moller, J. V. and Sheikh, M. I. (1982) *Pharmacol. Rev.* 34:315-358; Pritchard, J. B. and Miller, D. S. (1993) *Physiol. Rev.* 73:765-796; Ullrich, K. J. (1997) *J. Membr. Biol.* 158:95-107; Ullrich, K. J. and Rumrich, G. (1993) *Clin. Investig.* 71:843-848; Petzinger, E. (1994) *Rev. Physiol. Biochem. Pharmacol.* 123:47-211).

Summary of the Invention

[0008] The present invention is based, at least in part, on discovery of a human cDNA molecule which encodes a protein which is herein designated 65484. This protein catalyzes or facilitates transmembrane transport of charged

organic compounds, such as one or more of prostaglandins, thromboxanes, hexoses, disaccharides, hormones (e.g. insulin), peptides, neurotransmitters, cytokines, chemokines, and the like. The nucleotide sequence of a cDNA encoding 46584 is shown in SEQ ID NO:1, and the amino acid sequence of a 46584 polypeptide is shown in SEQ ID NO:2. In addition, the nucleotide sequence of the coding region is depicted in SEQ ID NO:3.

[0009] Accordingly, in one aspect, the invention features a nucleic acid molecule which encodes a 46584 protein or polypeptide, e.g., a biologically active portion of the 46584 protein. In a preferred embodiment, the isolated nucleic acid molecule encodes a polypeptide having the amino acid sequence of SEQ ID NO:2. In other embodiments, the invention provides isolated 46584 nucleic acid molecules having the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3. In still other embodiments, the invention provides nucleic acid molecules that are substantially identical (e.g., naturally occurring allelic variants) to the nucleotide sequence shown in SEQ ID NO: 1 or SEQ ID NO:3. In other embodiments, the invention provides a nucleic acid molecule which hybridizes under a stringent hybridization condition to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3, wherein the nucleic acid encodes a full length 46584 protein or an active fragment thereof.

[0010] In a related aspect, the invention further provides nucleic acid constructs which include a 46584 nucleic acid molecule described herein. In certain embodiments, the nucleic acid molecules of the invention are operatively linked to native or heterologous regulatory sequences. Also included are vectors and host cells containing the 46584 nucleic acid molecules of the invention e.g., vectors and host cells suitable for producing polypeptides.

[0011] In another related aspect, the invention provides nucleic acid fragments suitable as primers or hybridization probes for the detection of 46584-encoding nucleic acids.

[0012] In still another related aspect, isolated nucleic acid molecules that are antisense to a 46584 encoding nucleic acid molecule are provided.

[0013] In another aspect, the invention features 46584 polypeptides, and biologically active or antigenic fragments thereof that are useful, e.g., as reagents or targets in assays applicable to treatment and diagnosis of charged organic compound transporter-associated, or other 46584-associated disorders. In another embodiment, the invention provides 46584 polypeptides having a 46584 activity. Preferred polypeptides are 46584 proteins including at least one sugar transporter domain and, preferably, having a 46584 activity, e.g., a 46584 activity as described herein.

[0014] In other embodiments, the invention provides 46584 polypeptides, e.g., a 46584 polypeptide having the amino acid sequence shown in SEQ ID NO:2; an amino acid sequence that is substantially identical to the amino acid sequence shown in SEQ ID NO:2; or an amino acid sequence encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under a stringent hybridization condition to a nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3, wherein the nucleic acid encodes a full length 46584 protein or an active fragment thereof.

[0015] In a related aspect, the invention further provides nucleic acid constructs which include a 46584 nucleic acid molecule described herein.

[0016] In a related aspect, the invention provides 46584 polypeptides or fragments operatively linked to non-46584 polypeptides to form fusion proteins.

[0017] In another aspect, the invention features antibodies and antigen-binding fragments thereof, that react with, or more preferably specifically or selectively bind 46584 polypeptides.

[0018] In another aspect, the invention provides methods of screening for compounds that modulate the expression or activity of the 46584 polypeptides or nucleic acids.

[0019] In still another aspect, the invention provides a process for modulating 46584 polypeptide or nucleic acid expression or activity, e.g., using the compounds identified in the screens described herein. In certain embodiments, the methods involve treatment of conditions related to aberrant activity or expression of the 46584 polypeptides or nucleic acids, such as conditions or disorders involving aberrant or deficient charged organic compound transporter function or expression. Examples of such disorders include but are not limited to, a nervous system disorder, a cardiovascular disorder, a metabolic disorder, an immune, e.g., inflammatory, disorder, a kidney disorder, or a disorder involving aberrant cellular proliferation, differentiation, or migration.

[0020] The invention also provides assays for determining the activity of or the presence or absence of 46584 polypeptides or nucleic acid molecules in a biological sample, including for disease diagnosis.

[0021] In a further aspect, the invention provides assays for determining the presence or absence of a genetic alteration in a 46584 polypeptide or nucleic acid molecule, including for disease diagnosis.

[0022] In another aspect, the invention features a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence. At least one address of the plurality has a capture probe that recognizes a 46584 molecule. In one embodiment, the capture probe is a nucleic acid, e.g., a probe complementary to a 46584 nucleic acid sequence. In another embodiment, the capture probe is a polypeptide, e.g., an antibody specific for 46584 polypeptides. Also featured is a method of analyzing a sample by contacting the sample to the aforementioned array and detecting binding of the sample to the array.

[0023] Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Brief Description of the Drawings

[0024] Figure 1 depicts a hydropathy plot of human 46584. Relatively hydrophobic residues are shown above the dashed horizontal line, and relatively hydrophilic residues are below the dashed horizontal line. The cysteine residues (Cys) and N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace. The numbers corresponding to the amino acid sequence of human 46584 are indicated. Polypeptides of the invention include fragments which include: all or part of a hydrophobic sequence, e.g., a sequence above the dashed line; all or part of a hydrophilic sequence, e.g., a sequence below the dashed line; and a sequence which includes a cysteine, or a glycosylation site.

Detailed Description of the Invention

[0025] The human 46584 sequence (SEQ ID NO:1), which is approximately 6911 nucleotides long including untranslated regions, contains a predicted methionine-initiated coding sequence of about 2181 nucleotides, not including the termination codon (nucleotides indicated as coding of SEQ ID NO:1 and set forth in SEQ ID NO:3). The coding sequence encodes a 727 amino acid protein (SEQ ID NO:2).

[0026] Human 46584 contains the following regions or other structural features (for general information regarding PFAM identifiers, PS prefix and PF prefix domain identification numbers, refer to Sonnhammer *et al.* (1997) *Protein* 28:405-420 and <http://www.psc.edu/general/software/packages/pfam/pfam.html>):

a sugar transporter domain (PFAM Accession Number PF00083) located at about amino acid residues 138 to 727 of SEQ ID NO:2;

twelve transmembrane domains (predicted by MEMSAT, Jones *et al.* (1994) *Biochemistry* 33:3038-3049) at about amino acids 152 to 176, 192 to 212, 221 to 238, 249 to 267, 279 to 301, 321 to 340, 436 to 454, 581 to 603, 610 to 628, 635 to 651, 668 to 692, and 699 to 718 of SEQ ID NO:2;

one coiled coil structure (PSORT, <http://psort.nibb.ac.jp>) located at about amino acids 112 to 143 of SEQ ID NO:2;

one sugar transport protein signature sequence (Prosite PS00217) located at about amino acids 250 to 275 of SEQ ID NO:2;

five N-glycosylation sites (Prosite PS00001) located at about amino acids 480 to 483, 484 to 487, 534 to 537, 559 to 562, and 565 to 568 of SEQ ID NO:2;

two glycosaminoglycan attachment sites (Prosite PS00002) located at about amino acids 188 to 191, and 251 to 254 of SEQ ID NO:2;

three cAMP/cGMP-dependent protein kinase phosphorylation sites (Prosite PS00004) located at about amino acids 23 to 26, 217 to 220, and 666 to 669 of SEQ ID NO:2;

ten protein kinase C phosphorylation sites (Prosite PS00005) located at about amino acids 4 to 6, 26 to 28, 42 to 44, 125 to 127, 318 to 320, 343 to 345, 385 to 387, 433 to 435, 511 to 513, and 664 to 666 of SEQ ID NO:2;

eight casein kinase II phosphorylation sites (Prosite PS00006) located at about amino acids 4 to 7, 493 to 496, 511 to 514, 521 to 524, 536 to 539, 540 to 543, 545 to 548, and 572 to 575 of SEQ ID NO:2;

one tyrosine kinase phosphorylation site (Prosite PS00007) located at about amino acids 471 to 478 of SEQ ID NO:2;

twelve N-myristoylation sites (Prosite PS00008) located at about amino acids 74 to 79, 110 to 115, 200 to 205, 254 to 259, 289 to 294, 416 to 421, 568 to 573, 596 to 601, 630 to 635, 645 to 650, 690 to 695, and 711 to 716 of SEQ ID NO:2; and

one amidation site (Prosite PS00009) located at about amino acids 215 to 218.

[0027] The 46584 protein contains a significant number of structural characteristics in common with "charged organic compound transporters" or "COC transporters". As used herein, "COC transporters" are molecules capable of transporting charged organic compounds, including, e.g., members of the sugar transporter, organic anion transporter, and organic cation transporter families (e.g., SLC2, SLC21, and SLC22). The term "family" when referring to the protein and nucleic acid molecules of the invention means two or more proteins or nucleic acid molecules having a common structural domain or motif and having sufficient amino acid or nucleotide sequence homology as defined herein. Such family members can be naturally or non-naturally occurring and can be from either the same or different species. For example, a family can contain a first protein of human origin as well as other distinct proteins of human origin, or alternatively, can contain homologs of non-human origin, e.g., rat or mouse proteins. Members of a family also can have common functional characteristics.

[0028] As used herein, the term "sugar transporter," or "SLC2 family member" includes a protein or polypeptide which is capable of mediating facilitated diffusion, e.g., driven by the substrate concentration gradient, of a molecule, e.g. a monosaccharide (e.g. glucose, fructose, galactose or myo-inositol) across a membrane, e.g. a cell membrane (e.g. a membrane of a nerve cell, fat cell, muscle cell, or blood cell, such as an erythrocyte) or organelle (e.g. a mitochondrion) membrane. Sugar transporters play a role in or function in a variety of cellular processes, e.g., maintenance of sugar homeostasis and, typically, have sugar substrate specificity. Examples of sugar transporters include glucose transporters, fructose transporters, galactose transporters and yeast myo-inositol transporters.

[0029] The sugar transporter, or SLC2, family of proteins are characterized by twelve amphipathic (i.e. having hydrophilic or charged residue(s) along one face of an otherwise hydrophobic helix) transmembrane domains included within a sugar transporter domain, intracellular N- and C-termini, a large non-cytoplasmic hydrophilic loop between transmembrane domains one and two, a large cytoplasmic hydrophilic loop between transmembrane domains six and seven, and an oscillating pore mechanism of function (Barrett *et al.*, *supra*). Typically, the transmembrane domains anchor the transporter within a membrane and through coordinated allosteric movements, effect the transport function along their hydrophilic faces across the membrane, while contributing to the sugar type selectivity. The hydrophilic non-transmembrane loops between and beyond the transmembrane domains of the transporter determine the ion binding specificity and provide the ion binding sites, the trigger for the transport conformational change, and release activity for the transporter.

[0030] An alignment of the 46584 protein with *Rattus norvegicus* synaptic vesicle protein 2C (SV2C, assigned GenBank Accession Number AF060174, and set forth as SEQ ID NO:4) demonstrates about 97% sequence identity between the two sequences (as calculated in matblas from the blosum62.ij matrix). SV2C is an isoform of SV2, a protein found in all synaptic vesicles in all vertebrates and is also found on neuroendocrine granules (Lowe, *et al.* (1988) *J. Cell Biol.* 106:51-59). In addition to SV2C, there are two other isoforms, SV2A and SV2B. Sequence analysis of SV2 indicates the presence of of twelve transmembrane domains and significant homology to transport proteins, including bacterial and fungal proteins that cotransport protons with sugars, citrate, and drugs, as well as mammalian facilitative glucose transporters (Henderson and Maiden (1990) *Phil. Trans. R. Soc. Lond. Ser. B* 326:391-340; Maiden, *et al.* (1987) *Nature* 325:641-643).

[0031] A 46584 polypeptide can include a "sugar (and other) transporter domain" or regions homologous with a "sugar (and other) transporter domain". Sugar (or other) transport domains occur in a variety of proteins involved in transmembrane transport of sugars and other metabolites. Other proteins which comprise such a domain include human glucose transporters members of the SLC2 family, e.g., GLUT1, GLUT2, GLUT3, GLUT4, GLUT5, GLUT6, and GLUT7; *Escherichia coli* proteins AraE (arabinose-proton symporter); GalP (galactose-proton symporter); citrate-proton symport protein; KgtP (α -ketoglutarate permease); ProP (proline/betaine transporter); and Xyle (xylose-proton symporter); *Escherichia coli* hypothetical proteins YabE, YdjE, and YhjE; *Klebsiella pneumoniae* citrate-proton symport protein; *Zymomonas mobilis* glucose facilitated diffusion protein; yeast high and low affinity glucose transport proteins (SNF3 and HXT1 through HXT14); yeast galactose transporter; yeast maltose permease; yeast myo-inositol transporter; yeast carboxylic acid transporter homolog JEN1; yeast hypothetical proteins YBR241c, YCR98c, and YFL040w; *Kluyveromyces fragilis* lactose permease; *Neurospora crassa* quinate transporter; *Emericella nidulans* quinate permease; *Chlorella* hexose carrier; *Arabidopsis thaliana* glucose transporter; spinach sucrose transporter; *Leishmania donovani* transporters D1 and D2; *Leishmania enriettii* probable transport protein LTP; *Caenorhabditis elegans* hypothetical protein ZK637.1; *Haemophilus influenzae* hypothetical proteins HI0281 and HI0418; and *Bacillus subtilis* hypothetical proteins YxbC and YxdF. Occurrence of a sugar (or other) transport domain in 46584 polypeptide indicates that the polypeptide is involved in transmembrane transport of one or more compounds, most likely a compound having a molecular weight on the order of a hexose or greater (i.e. having a molecular weight greater than about 180). Examples of such compounds include prostaglandins, thromboxanes, hexoses, disaccharides, hormones (e.g. insulin), peptides, neurotransmitters, cytokines, chemokines, and the like. 46584 thus mediates one or more of facilitated diffusion and symport or antiport (e.g. involving co-transport of a proton, a sodium ion, a potassium ion, or another physiological ion).

[0032] As used interchangeably herein, the term "sugar transporter domain" or "sugar (and other) transporter domain" includes an amino acid sequence of about 450 to 650 amino acid residues in length and having a bit score for the alignment of the sequence to the sugar transporter domain (HMM) of at least 25, preferably at least 35, more preferably at least 38. Preferably a sugar transporter domain mediates facilitated diffusion of molecules, e.g. monosaccharides (e.g., glucose, fructose, galactose, or myo-inositol) across a membrane. Preferably, a sugar transporter domain includes at least about 525 to 625 amino acids, more preferably about 570 to 610 amino acid residues, or about 585 to 595 amino acids and has a bit score for the alignment of the sequence to the sugar transporter domain (HMM) of at least 30, 35, 38 or greater, or an E-value of no more than 1.0×10^{-3} , 1.0×10^{-6} , 1.0×10^{-8} , or less. A consensus sequence for a sugar transporter domain has been assigned the PFAM Accession Number PF00083 (<http://genome.wustl.edu/Pfam/>.html).

[0033] An alignment of the sugar transporter domain (amino acids 138 to 727 of SEQ ID NO:2) of human 46584 with Pfam Accession Number PF00083 (SEQ ID NO:5) derived from a hidden Markov model has a bit score of 39.9 and

e-value of 3.6e-08.

[0034] A sugar transporter domain can also be found in family members of other classes of transporters, *e.g.*, the organic anion/prostaglandin family (SLC21) and the organic cation family (SLC22). The members of the SLC21 family include the organic anion transporting proteins OATP1, OATP2, OATP3, OATP3, and OAT-K2, as well as the prostaglandin transporter PGT. The members of the SLC22 family include the electrogenic, pH-independent organic cation transporters OCT1, OCT2, and OCT3, the pH-dependent antiporter OCTN1, the Na⁺-dependent carnitine transporter OCTN2, and the organic anion transporters OAT1, OAT2, and OAT3.

[0035] As used herein, the terms "organic anion/prostaglandin family", "organic cation family", "SLC21", or "SLC22" include secondary active transport proteins. Secondary active transporters couple the active transport of one molecule, *e.g.*, an ion, *e.g.*, an organic ion (*e.g.*, an organic anion or a cation, a prostaglandin, a steroidal compound, a bile acid, a drug, a neurotransmitter, a sulfated lipophilic metabolite, a glucuronidated lipophilic metabolite, a polyamine, a carnitine, or a choline) against its concentration gradient to the energy gained by concomitant transport of a second molecule, *e.g.*, another ion (*e.g.*, a proton or a sodium ion) with its concentration gradient.

[0036] The SLC21 and SLC22 families of proteins of proteins are characterized by at least one, two, three, four, five, six, seven, eight, nine, ten, eleven, preferably twelve transmembrane domains. Typically, the hydrophobic transmembrane domains anchor the transporter within a cell or organelle membrane and through coordinated allosteric movements, effect the transport function across the membrane. The non-transmembrane loops between and beyond the transmembrane domains of the transporter determine the ion binding specificity and provide the ion binding and release activity for the transporter.

[0037] In a preferred embodiment, a 46584 polypeptide or protein has a "sugar transporter domain" or a region which includes at least about 525 to 625 amino acids, more preferably about 570 to 610 amino acid residues, or about 585 to 595 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a "sugar transporter domain," *e.g.*, the sugar transporter domain of human 46584 (*e.g.*, residues 138 to 727 of SEQ ID NO:2).

[0038] To identify the presence of a "sugar transporter domain" domain in a 46584 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be searched against the Pfam database of HMMs (*e.g.*, the Pfam database, release 2.1) using the default parameters (http://www.sanger.ac.uk/Software/Pfam/HMM_search). For example, the hmmsf program, which is available as part of the HMMER package of search programs, is a family specific default program for MILPAT0063 and a score of 15 is the default threshold score for determining a hit. Alternatively, the threshold score for determining a hit can be lowered (*e.g.*, to 8 bits). A description of the Pfam database can be found in Sonhammer *et al.* (1997) *Proteins* 28: 405-420 and a detailed description of HMMs can be found, for example, in Gribskov *et al.* (1990) *Meth. Enzymol.* 183: 146-159; Gribskov *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:4355-4358; Krogh *et al.* (1994) *J. Mol. Biol.* 235: 1501-1531; and Stultz *et al.* (1993) *Protein Sci.* 2:305-314, the contents of which are incorporated herein by reference. A search was performed against the HMM database resulting in the identification of a "sugar transporter domain" domain in the amino acid sequence of human 46584 at about residues 138 to 727 of SEQ ID NO:2.

[0039] A COC transporter can include one or more signature sequences, *e.g.*, a sugar transport protein signature sequence, or sequences homologous thereto, in the protein or encoded by the corresponding nucleic acid sequence. As used herein, a "sugar transport protein signature sequence", or a sequence homologous thereto, is a conserved signature sequence characteristic of sugar transporters. The consensus sequence for a sugar transport protein signature sequence is [LIVMF]-x-G-[LIVMFA]-x(2)-G-x(8)-[LIFY]-x(2)-[EQ]-x(6)-[RK] (SEQ ID NO:6) and has been assigned Prosite Accession No. PS00217 (<http://www.expasy.ch/prosite/>). In the above consensus signature sequence, and other signature sequences or motifs described herein, the standard IUPAC one-letter code for the amino acids is used. Each element in the pattern is separated by a dash (-); square brackets ([]) indicate the particular residues that are accepted at that position; X indicates that any residue is accepted at that position; and numbers in parentheses (()) indicate the number of residues represented by the accompanying amino acid. This consensus sequence is based on a number of conserved residues which are located at the end of the fourth transmembrane segment and in the short loop region between the fourth and fifth segments.

[0040] In a preferred embodiment, the sugar transport protein signature sequence is located within the sugar transporter domain at about the fourth transmembrane and third cytoplasmic domains of the human 46584 polypeptide and which corresponds to about amino acids 250 to 275 of SEQ ID NO:2.

[0041] A COC transporter can further include a coiled coil structure. Coiled coil structures are supercoiled helical domains responsible for the oligomerization of proteins. There is a characteristic heptad repeat (h-x-x-h-x-x-x)n in the coiled coil structures, where h represents hydrophobic residues (Beck and Brodsky (1998) *J. Struct. Biol.* 122:17-29). Coiled coil structures are found in a wide variety of proteins, including cytoskeletal, nuclear, muscle, cell surface, extracellular, plasma, bacterial, and viral proteins.

[0042] In a preferred embodiment, the coiled coil structure is located within the N-terminal cytoplasmic domain of the human 46584 polypeptide and which corresponds to about amino acids 112 to 143 of SEQ ID NO:2.

[0043] A 46584 polypeptide can include at least one, two, three, four, five, six, preferably seven, eight, nine, ten,

eleven, more preferably twelve "transmembrane domains" or regions homologous with "transmembrane domains". As used herein, the term "transmembrane domain" includes an amino acid sequence of about 10 to 40 amino acid residues in length and spans the plasma membrane. Transmembrane domains are rich in hydrophobic residues, e.g., at least 50%, 60%, 70%, 80%, 90%, 95% or more of the amino acids of a transmembrane domain are hydrophobic, e.g., leucines, isoleucines, tyrosines, or tryptophans. Transmembrane domains typically have alpha-helical structures and are described in, for example, Zagotta *et al.*, (1996) *Annual Rev. Neurosci.* 19:235-263, the contents of which are incorporated herein by reference. The transmembrane domains of human 46584 are located at about residues 152 to 176, 192 to 212, 221 to 238, 249 to 267, 279 to 301, 321 to 340, 436 to 454, 581 to 603, 610 to 628, 635 to 651, 668 to 692, and 699 to 718 of SEQ ID NO:2.

[0044] In a preferred embodiment, a 46584 polypeptide or protein has at least one, two, three, four, five, six, preferably seven, eight, nine, ten, eleven, more preferably twelve "transmembrane domain" or regions which include at least about 12 to 35, more preferably about 14 to 30, or 15 to 27 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a "transmembrane domain," e.g., the transmembrane domains of human 46584 (e.g., residues 152 to 176, 192 to 212, 221 to 238, 249 to 267, 279 to 301, 321 to 340, 436 to 454, 581 to 603, 610 to 628, 635 to 651, 668 to 692, and 699 to 718 of SEQ ID NO:2). The transmembrane domains of human 46584 are visualized in the hydropathy plot shown in Figure 1 as regions of about 15 to 25 amino acids where the hydropathy trace is mostly above the horizontal line.

[0045] To identify the presence of a "transmembrane" domain in a 46584 protein sequence, and make the determination that a polypeptide or protein of interest has a particular profile, the amino acid sequence of the protein can be analyzed by a transmembrane prediction method that predicts the secondary structure and topology of integral membrane proteins based on the recognition of topological models (MEMSAT, Jones *et al.*, (1994) *Biochemistry* 33: 3038-3049).

[0046] A 46584 polypeptide can include at least one, two three, four, five, six, seven, eight, preferably nine, ten, eleven, twelve, more preferably thirteen "non-transmembrane regions." As used herein, the term "non-transmembrane region" includes an amino acid sequence not identified as a transmembrane domain. The non-transmembrane regions in 46584 are located at about amino acids 1 to 151, 177 to 191, 213 to 220, 239 to 248, 268 to 278, 302 to 320, 341 to 435, 455 to 580, 604 to 609, 629 to 634, 652 to 667, 693 to 698, and 719 to 727 of SEQ ID NO:2.

[0047] The non-transmembrane regions of 46584 include at least one, two, three, preferably four, five, six, more preferably seven cytoplasmic regions. When located at the N-terminus, the cytoplasmic region is referred to herein as the "N-terminal cytoplasmic domain." As used herein, an "N-terminal cytoplasmic domain" includes an amino acid sequence having about 1 to 180, preferably about 1 to 170, more preferably about 1 to 160, or even more preferably about 1 to 151 amino acid residues in length, is located inside of a cell or within the cytoplasm of a cell. The C-terminal amino acid residue of an "N-terminal cytoplasmic domain" is adjacent to an N-terminal amino acid residue of a transmembrane domain in a 46584 protein. For example, an N-terminal cytoplasmic domain is located at about amino acid residues 1 to 151 of SEQ ID NO:2.

[0048] In a preferred embodiment, a 46584 polypeptide or protein has an N-terminal cytoplasmic domain or a region which includes about 1 to 180, preferably about 1 to 170, more preferably about 1 to 160, or even more preferably about 1 to 151 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with an "N-terminal cytoplasmic domain," e.g., the N-terminal cytoplasmic domain of human 46584 (e.g., residues 1 to 151 of SEQ ID NO:2).

[0049] In another embodiment, a 46584 cytoplasmic region includes at least one, two, three, preferably four, more preferably five cytoplasmic loops. As used herein, the term "loop" includes an amino acid sequence which is not included within a phospholipid membrane, having a length of at least about 5, preferably about 6 to 140, more preferably about 6 to 130 amino acid residues, and has an amino acid sequence that connects two transmembrane domains within a protein or polypeptide. Accordingly, the N-terminal amino acid of a loop is adjacent to a C-terminal amino acid of a transmembrane domain in a 46584 molecule, and the C-terminal amino acid of a loop is adjacent to an N-terminal amino acid of a transmembrane domain in a 46584 molecule. As used herein, a "cytoplasmic loop" includes a loop located inside of a cell or within the cytoplasm of a cell. For example, a "cytoplasmic loop" can be found at about amino acid residues 213 to 220, 268 to 278, 341 to 435, 604 to 609, and 652 to 667 of SEQ ID NO:2.

[0050] In a preferred embodiment, a 46584 polypeptide or protein has a cytoplasmic loop or a region which includes at least about 4, preferably about 5 to 100, and more preferably about 6 to 95 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a cytoplasmic loop," e.g., a cytoplasmic loop of human 46584 (e.g., residues 213 to 220, 268 to 278, 341 to 435, 604 to 609, and 652 to 667 of SEQ ID NO:2).

[0051] In another embodiment, a 46584 non-transmembrane region includes at least one, one, two, three, preferably four, five, more preferably six non-cytoplasmic loops. As used herein, a "non-cytoplasmic loop" includes a loop located outside of a cell or within an intracellular organelle. Non-cytoplasmic loops include extracellular domains (*i.e.*, outside of the cell) and intracellular domains (*i.e.*, within the cell). When referring to membrane-bound proteins found in intracellular organelles (e.g., mitochondria, endoplasmic reticulum, peroxisomes microsomes, vesicles, endosomes, and

lysosomes), non-cytoplasmic loops include those domains of the protein that reside in the lumen of the organelle or the matrix or the intermembrane space. For example, a "non-cytoplasmic loop" can be found at about amino acid residues 177 to 191, 239 to 248, 302 to 320, 455 to 580, 629 to 634, and 693 to 698 of SEQ ID NO:2.

[0052] In a preferred embodiment, a 46584 polypeptide or protein has at least one non-cytoplasmic loop or a region which includes at least about 4, preferably about 5 to 135, more preferably about 6 to 126 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a "non-cytoplasmic loop," e.g., at least one non-cytoplasmic loop of human 46584 (e.g., residues 177 to 191, 239 to 248, 302 to 320, 455 to 580, 629 to 634, and 693 to 698 of SEQ ID NO:2).

[0053] In another embodiment, a cytoplasmic region of a 46584 protein can include the C-terminus and can be a "C-terminal cytoplasmic domain," also referred to herein as a "C-terminal cytoplasmic tail." As used herein, a "C-terminal cytoplasmic domain" includes an amino acid sequence having a length of at least about 5, preferably about 6 to 12, more preferably about 8 to 10 amino acid residues, is located inside of a cell or within the cytoplasm of a cell. The N-terminal amino acid residue of a "C-terminal cytoplasmic domain" is adjacent to a C-terminal amino acid residue of a transmembrane domain in a 46584 protein. For example, a C-terminal cytoplasmic domain is located at about amino acid residues 719 to 727 of SEQ ID NO:2.

[0054] In a preferred embodiment, a 46584 polypeptide or protein has a C-terminal cytoplasmic domain or a region which includes at least about 5, preferably about 6 to 12, more preferably about 8 to 10 amino acid residues and has at least about 60%, 70%, 80%, 90%, 95%, 99%, or 100% homology with a C-terminal cytoplasmic domain," e.g., the C-terminal cytoplasmic domain of human 46584 (e.g., residues 719 to 727 of SEQ ID NO:2).

[0055] A 46584 family member can include the following structural features: one sugar transporter domain, twelve transmembrane domains, and thirteen non-transmembrane domains; at least one sugar transport protein signature sequence, located preferably within the sugar transporter domain; at least one coiled coil structure; at least one, two, three, preferably four, more preferably five N-glycosylation sites (Prosite PS00001); one, preferably two glycosaminoglycan attachment sites (Prosite PS00002); one, preferably two, more preferably three cAMP/cGMP-dependent protein kinase phosphorylation sites (Prosite PS00004); one, two, three, four, five, preferably six, seven, eight, nine, more preferably ten protein kinase C phosphorylation sites (Prosite PS00005); one, two, three, four, preferably five, six, seven, more preferably eight casein kinase II phosphorylation sites (Prosite PS00006); one tyrosine kinase phosphorylation site (Prosite PS00007); one, two, three, four, five, six, preferably seven, eight, nine, ten, eleven, more preferably twelve N-myristoylation sites (Prosite PS00008); and one amidation site (Prosite PS00009).

[0056] As the 46584 polypeptides of the invention can modulate 46584-mediated activities, they can be useful for developing novel diagnostic and therapeutic agents for COC transporter-associated or other 46584-associated disorders, as described below.

[0057] As used herein, a "COC transporter-mediated activity" includes an activity which involves transport of a molecule, e.g., a charged organic compound, e.g., a monosaccharide (e.g. glucose, fructose, galactose or myo-inositol), an organic anion (e.g., bilirubin or sulfobromophthalein, taurocholate, or bile salts), a prostaglandin (e.g., PGE2 or PGF2 α), or an organic cation (e.g., choline, N-methylnicotinamide, polyamines, or monoamine neurotransmitters) across a membrane, e.g. a cell membrane (e.g. a membrane of a nerve cell, fat cell, muscle cell, or blood cell, such as an erythrocyte) or organelle (e.g. a mitochondrion) membrane.

[0058] Sugar transporters of the SLC2 family play important roles in sugar homeostasis, i.e., in making monosaccharides available to cells to use as an energy source. Besides a general role in metabolism, this sugar homeostatic role of sugar transporters is associated with systems with specific or high energy demands, e.g. with the neurological and cardiovascular systems. As a result, glucose transporters are being investigated in relation to infantile seizures (Klepper *et al.* (1999) *Neurochem. Res.* 24:587-94) and coronary artery disease (Young *et al.* (1999) *Am. J. Cardiol.* 83:25H-30H).

[0059] The organic ion transport function of the SLC21 and SLC22 families is important for many physiological processes. At the cellular level, aberrant or deficient organic ion transporter activity can detrimentally affect functions such as cellular proliferation, growth, differentiation, or migration. At the tissue level, aberrant or deficient organic ion transporter activity can detrimentally affect inter- or intra-cellular communication; or musculoskeletal function. At the organ level, aberrant or deficient organic ion transporter activity can detrimentally affect kidney, liver or cardiac function. At the organism level, aberrant or deficient organic ion transporter activity can detrimentally affect systemic responses, such as nervous system responses, hormonal responses (e.g., insulin response), or immune responses; and protection of cells from toxic compounds (e.g., carcinogens, toxins, or mutagens).

[0060] As used herein, a "46584 activity", "biological activity of 46584" or "functional activity of 46584", refers to an activity exerted by a 46584 protein, polypeptide or nucleic acid molecule on e.g., a 46584-responsive cell or on a 46584 substrate, e.g., a protein substrate, as determined *in vivo* or *in vitro*. In one embodiment, a 46584 activity is a direct activity, such as an association with a 46584 target molecule. A "target molecule" or "binding partner" is a molecule with which a 46584 protein binds or interacts in nature. In an exemplary embodiment, 46584 is transporter, e.g., a charged organic compound transporter, and thus binds to or interacts in nature with a molecule, e.g., a monosaccharide

(e.g. glucose, fructose, galactose or myo-inositol), an organic anion (e.g., bilirubin or sulfobromophthalein, taurocholate, or bile salts), a prostaglandin (e.g., PGE2 or PGF2 α), or an organic cation (e.g., choline, N-methylnicotinamide, polyamines, or monoamine neurotransmitters).

[0061] A 46584 activity can also be an indirect activity, e.g., a cellular signaling activity mediated by interaction of the 46584 protein with a 46584 receptor.

[0062] Based on the above-described sequence structures and similarities to molecules of known function, the 46584 molecules of the present invention can have similar biological activities as molecules with charged organic compound transporter function, e.g., members of the SLC2, SLC21, or SLC22 family. For example, the 46584 proteins of the present invention can have one or more of the following activities: (1) the ability to reside within a membrane, e.g., a cell membrane (e.g., a nerve cell membrane, fat cell membrane, muscle cell membrane, and/or blood cell membrane, such as an erythrocyte membrane) or organelle (e.g., mitochondrion) membrane; (2) the ability to interact with a substrate or target molecule, e.g., a monosaccharide, (e.g., glucose, fructose, galactose or myo-inositol) or an organic ion (e.g., an organic anion, an organic cation, a prostaglandin, a steroidal compound, a bile acid, a drug, a neurotransmitter, a sulfated lipophilic metabolite, a glucuronidated lipophilic metabolite, a polyamine, a carnitine, or a choline) across a membrane; (3) the ability to transport the substrate or target molecule across the membrane; (4) the ability to transport a second substrate or target molecule, e.g., an ion, (e.g., a bicarbonate ion or a dicarboxylate ion), across a membrane; (5) the ability to interact with and/or modulate a second non-transporter protein; (6) the ability to modulate cellular signaling and/or gene transcription (e.g., either directly or indirectly); (7) the ability to protect cells and/or tissues from organic ions; (8) the ability to protect cells and/or tissues from organic anions; (9) the ability to modulate hormonal responses, e.g., insulin and/or glucagon secretion; (10) the ability to modulate metabolism; (11) the ability to modulate excretion; and (12) the ability to modulate sugar homeostasis in a cell.

[0063] Thus, the 46584 molecules can act as novel diagnostic targets and therapeutic agents for prognosticating, diagnosing, preventing, inhibiting, alleviating, or curing one or more COC transporter disorders. As used herein, a "COC transporter disorder" or "COC transporter-associated disorder" includes a disorder, disease, or condition which is caused by, characterized by, or associated with a misregulation, e.g., an aberrant or deficient (e.g., downregulated or upregulated), of a COC transporter mediated activity. COC transporter-associated disorders can result in an aberrant or deficient intracellular sugar or organic ion organic levels, inter- or intra-cellular signaling, and/or hormonal response. Examples of COC transporter-associated disorders include disorders associated with aberrant sugar homeostasis, such as obesity, anorexia, type-1 diabetes, type-2 diabetes, hypoglycemia, glycogen storage disease (Von Gierke disease), type I glycogenosis, bipolar disorder, seasonal affective disorder, and cluster B personality disorders. Other COC transporter-associated disorders include disorders associated with carnitine deficiency, such as muscular dystrophy, polymyositis, Fancon's syndrome, and organic acidemia.

[0064] Other activities, as described below, include the ability to modulate function, survival, morphology, proliferation and/or differentiation of, and oligopeptide uptake by cells of tissues in which 46584 molecules are expressed. Thus, the 46584 molecules can act as novel diagnostic targets and therapeutic agents for controlling disorders involving aberrant activities of these cells.

[0065] Additional examples of disorders associated with aberrant or deficient COC transporter protein function or expression, e.g., COC transporter-associated or other 46584-associated disorders, include, but are not limited to, cellular proliferative and/or differentiative disorders, neurological disorders, bone metabolism, immune, e.g., inflammatory, disorders, cardiovascular disorders, endothelial cell disorders, liver disorders, viral diseases, pain or metabolic disorders.

[0066] Examples of cellular proliferative and/or differentiative disorders include cancer, e.g., carcinoma, sarcoma, metastatic disorders or hematopoietic neoplastic disorders, e.g., leukemias. A metastatic tumor can arise from a multitude of primary tumor types, including but not limited to those of prostate, colon, lung, breast and liver origin.

[0067] As used herein, the term "cancer" (also used interchangeably with the terms, "hyperproliferative" and "neoplastic") refers to cells having the capacity for autonomous growth, i.e., an abnormal state or condition characterized by rapidly proliferating cell growth. Cancerous disease states may be categorized as pathologic, i.e., characterizing or constituting a disease state, e.g., malignant tumor growth, or may be categorized as non-pathologic, i.e., a deviation from normal but not associated with a disease state, e.g., cell proliferation associated with wound repair. The term is meant to include all types of cancerous growths or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. The term "cancer" includes malignancies of the various organ systems, such as those affecting lung, breast, thyroid, lymphoid, gastrointestinal, and genito-urinary tract, as well as adenocarcinomas which include malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer and/or testicular tumors, non-small cell carcinoma of the lung, cancer of the small intestine and cancer of the esophagus. The term "carcinoma" is art recognized and refers to malignancies of epithelial or endocrine tissues including respiratory system carcinomas, gastrointestinal system carcinomas, genitourinary system carcinomas, testicular carcinomas, breast carcinomas, prostatic carcinomas, endocrine system carcinomas, and melanomas. Exemplary carcinomas include those forming from tissue of the cervix, lung, prostate, breast, head and

neck, colon and ovary. The term "carcinoma" also includes carcinosarcomas, e.g., which include malignant tumors composed of carcinomatous and sarcomatous tissues. An "adenocarcinoma" refers to a carcinoma derived from glandular tissue or in which the tumor cells form recognizable glandular structures. The term "sarcoma" is art recognized and refers to malignant tumors of mesenchymal derivation.

[0068] The 46584 molecules of the invention can be used to monitor, treat and/or diagnose a variety of proliferative disorders. Such disorders include hematopoietic neoplastic disorders. As used herein, the term "hematopoietic neoplastic disorders" includes diseases involving hyperplastic/neoplastic cells of hematopoietic origin, e.g., arising from myeloid, lymphoid or erythroid lineages, or precursor cells thereof. Preferably, the diseases arise from poorly differentiated acute leukemias, e.g., erythroblastic leukemia and acute megakaryoblastic leukemia. Additional exemplary myeloid disorders include, but are not limited to, acute promyeloid leukemia (APML), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML) (reviewed in Vaickus (1991) *Crit Rev. in Oncol./Hematol.* 11:267-97); lymphoid malignancies include, but are not limited to acute lymphoblastic leukemia (ALL) which includes B-lineage ALL and T-lineage ALL, chronic lymphocytic leukemia (CLL), prolymphocytic leukemia (PLL), hairy cell leukemia (HLL) and Waldenstrom's macroglobulinemia (WM). Additional forms of malignant lymphomas include, but are not limited to non-Hodgkin lymphoma and variants thereof, peripheral T cell lymphomas, adult T cell leukemia/lymphoma (ATL), cutaneous T-cell lymphoma (CTCL), large granular lymphocytic leukemia (LGF), Hodgkin's disease and Reed-Sternberg disease.

[0069] The 46584 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of central nervous system (CNS) or neurological disorders. Examples of CNS disorders include cognitive and neurodegenerative disorders, examples of which include, but are not limited to, autonomic function disorders such as hypertension and sleep disorders, and neuropsychiatric disorders, such as depression, schizophrenia, schizoaffective disorder, Korsakoff's psychosis, anxiety disorders, or phobic disorders; learning or memory disorders, e.g., amnesia or age-related memory loss, attention deficit disorder, dysthymic disorder, major depressive disorder, mania, obsessive-compulsive disorder, psychoactive substance use disorders, anxiety, phobias, panic disorder, as well as bipolar affective disorder, e.g., severe bipolar affective (mood) disorder (BP-1), and bipolar affective neurological disorders, e.g., migraine and obesity. Such neurological disorders include, for example, disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural abscess, chronic bacterial meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme disease), viral meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicella-zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer's disease and Pick's disease, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson's disease (paralysis agitans) and other Lewy diffuse body diseases, progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy, including striatonigral degeneration, Shy-Drager syndrome, and olivopontocerebellar atrophy, and Huntington's disease, senile dementia, Gilles de la Tourette's syndrome, epilepsy, and Jakob-Creutzfeldt disease; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telanglectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as leukodystrophies, including Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamine (vitamin B₁) deficiency and vitamin B₁₂ deficiency, neurologic sequelae of metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephatopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary

(diffuse) astrocytoma and glioblastoma multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma, oligodendroglioma, and ependymoma and related paraventricular mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors, including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromatosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease. Further CNS-related disorders include, for example, those listed in the American Psychiatric Association's Diagnostic and Statistical manual of Mental Disorders (DSM), the most current version of which is incorporated herein by reference in its entirety.

[0070] Aberrant expression and/or activity of 46584 molecules can mediate disorders associated with bone metabolism. "Bone metabolism" refers to direct or indirect effects in the formation or degeneration of bone structures, e.g., bone formation, bone resorption, etc., which can ultimately affect the concentrations in serum of calcium and phosphate. This term also includes activities mediated by 46584 molecules in bone cells, e.g. osteoclasts and osteoblasts, that can in turn result in bone formation and degeneration. For example, 46584 molecules can support different activities of bone resorbing osteoclasts such as the stimulation of differentiation of monocytes and mononuclear phagocytes into osteoclasts. Accordingly, 46584 molecules that modulate the production of bone cells can influence bone formation and degeneration, and thus can be used to treat bone disorders. Examples of such disorders include, but are not limited to, osteoporosis, osteodystrophy, osteomalacia, rickets, osteitis fibrosa cystica, renal osteodystrophy, osteosclerosis, anti-convulsant treatment, osteopenia, fibrogenesis-imperfecta ossium, secondary hyperparathyroidism, hypoparathyroidism, hyperparathyroidism, cirrhosis, obstructive jaundice, drug induced metabolism, medullary carcinoma, chronic renal disease, rickets, sarcoidosis, glucocorticoid antagonism, malabsorption syndrome, steatorrhea, tropical sprue, idiopathic hypercalcemia and milk fever.

[0071] The 46584 nucleic acid and protein of the invention can be used to treat and/or diagnose a variety of immune, e.g., inflammatory, (e.g., respiratory inflammatory) disorders. Examples of immune disorders or diseases include, but are not limited to, autoimmune diseases (including, for example, diabetes mellitus, arthritis (including rheumatoid arthritis, juvenile rheumatoid arthritis, osteoarthritis, psoriatic arthritis), multiple sclerosis, encephalomyelitis, myasthenia gravis, systemic lupus erythematosus, autoimmune thyroiditis, dermatitis (including atopic dermatitis and eczematous dermatitis), psoriasis, Sjögren's Syndrome, inflammatory bowel disease, e.g. Crohn's disease and ulcerative colitis, aphthous ulcer, iritis, conjunctivitis, keratoconjunctivitis, asthma, allergic asthma, chronic obstructive pulmonary disease, cutaneous lupus erythematosus, scleroderma, vaginitis, proctitis, drug eruptions, leprosy reversal reactions, erythema nodosum leprosum, autoimmune uveitis, allergic encephalomyelitis, acute necrotizing hemorrhagic encephalopathy, idiopathic bilateral progressive sensorineural hearing loss, aplastic anemia, pure red cell anemia, idiopathic thrombocytopenia, polychondritis, Wegener's granulomatosis, chronic active hepatitis, Stevens-Johnson syndrome, idiopathic sprue, lichen planus, Graves' disease, sarcoidosis, primary biliary cirrhosis, uveitis posterior, and interstitial lung fibrosis), graft-versus-host disease, cases of transplantation, and allergy such as, atopic allergy.

[0072] As used herein, disorders involving the heart, or "cardiovascular disease" or a "cardiovascular disorder" includes a disease or disorder which affects the cardiovascular system, e.g., the heart, the blood vessels, and/or the blood. A cardiovascular disorder can be caused by an imbalance in arterial pressure, a malfunction of the heart, or an occlusion of a blood vessel, e.g., by a thrombus. A cardiovascular disorder includes, but is not limited to disorders such as arteriosclerosis, atherosclerosis, cardiac hypertrophy, ischemia reperfusion injury, restenosis, arterial inflammation, vascular wall remodeling, ventricular remodeling, rapid ventricular pacing, coronary microembolism, tachycardia, bradycardia, pressure overload, aortic bending, coronary artery ligation, vascular heart disease, valvular disease, including but not limited to, valvular degeneration caused by calcification, rheumatic heart disease, endocarditis, or complications of artificial valves; atrial fibrillation, long-QT syndrome, congestive heart failure, sinus node dysfunction, angina, heart failure, hypertension, atrial fibrillation, atrial flutter, pericardial disease, including but not limited to, pericardial effusion and pericarditis; cardiomyopathies, e.g., dilated cardiomyopathy or idiopathic cardiomyopathy, myocardial infarction, coronary artery disease, coronary artery spasm, ischemic disease, arrhythmia, sudden cardiac death, and cardiovascular developmental disorders (e.g., arteriovenous malformations, arteriovenous fistulae, raynaud's syndrome, neurogenic thoracic outlet syndrome, causalgia/reflex sympathetic dystrophy, hemangioma, aneurysm, cavernous angioma, aortic valve stenosis, atrial septal defects, atrioventricular canal, coarctation of the aorta, ebsteins anomaly, hypoplastic left heart syndrome, interruption of the aortic arch, mitral valve prolapse, ductus arteriosus, patent foramen ovale, partial anomalous pulmonary venous return, pulmonary atresia with ventricular septal defect, pulmonary atresia without ventricular septal defect, persistence of the fetal circulation, pulmonary valve stenosis, single ventricle, total anomalous pulmonary venous return, transposition of the great vessels, tricuspid atresia, truncus arteriosus, ventricular septal defects). A cardiovascular disease or disorder also can include an endothelial cell disorder.

[0073] As used herein, an "endothelial cell disorder" includes a disorder characterized by aberrant, unregulated, or unwanted endothelial cell activity, e.g., proliferation, migration, angiogenesis, or vascularization; or aberrant expression

of cell surface adhesion molecules or genes associated with angiogenesis, e.g., TIE-2, FLT and FLK. Endothelial cell disorders include tumorigenesis, tumor metastasis, psoriasis, diabetic retinopathy, endometriosis, Grave's disease, ischemic disease (e.g., atherosclerosis), and chronic inflammatory diseases (e.g., rheumatoid arthritis).

[0074] Disorders which can be treated or diagnosed by methods described herein include, but are not limited to, disorders associated with an accumulation in the liver of fibrous tissue, such as that resulting from an imbalance between production and degradation of the extracellular matrix accompanied by the collapse and condensation of pre-existing fibers.

The methods described herein can be used to diagnose or treat hepatocellular necrosis or injury induced by a wide variety of agents including processes which disturb homeostasis, such as an inflammatory process, tissue damage resulting from toxic injury or altered hepatic blood flow, and infections (e.g., bacterial, viral and parasitic). For example, the methods can be used for the early detection of hepatic injury, such as portal hypertension or hepatic fibrosis. In addition, the methods can be employed to detect liver fibrosis attributed to inborn errors of metabolism, for example, fibrosis resulting from a storage disorder such as Gaucher's disease (lipid abnormalities) or a glycogen storage disease, A1-antitrypsin deficiency; a disorder mediating the accumulation (e.g., storage) of an exogenous substance, for example, hemochromatosis (iron-overload syndrome) and copper storage diseases (Wilson's disease), disorders resulting in the accumulation of a toxic metabolite (e.g., tyrosinemia, fructosemia and galactosemia) and peroxisomal disorders (e.g., Zellweger syndrome). Additionally, the methods described herein can be used for the early detection and treatment of liver injury associated with the administration of various chemicals or drugs, such as for example, methotrexate, isoniazid, oxyphenisatin, methyl dopa, chlorpromazine, tolbutamide or alcohol, or which represents a hepatic manifestation of a vascular disorder such as obstruction of either the intrahepatic or extrahepatic bile flow or an alteration in hepatic circulation resulting, for example, from chronic heart failure, veno-occlusive disease, portal vein thrombosis or Budd-Chiari syndrome.

[0075] Additionally, 46584 molecules can play an important role in the etiology of certain viral diseases, including but not limited to Hepatitis B, Hepatitis C and Herpes Simplex Virus (HSV). Modulators of 46584 activity could be used to control viral diseases. The modulators can be used in the treatment and/or diagnosis of viral infected tissue or virus-associated tissue fibrosis, especially liver and liver fibrosis. Also, 46584 modulators can be used in the treatment and/or diagnosis of virus-associated carcinoma, especially hepatocellular cancer.

[0076] Additionally, 46584 can play an important role in the regulation of metabolism or pain disorders. Diseases of metabolic imbalance include, but are not limited to, obesity, anorexia nervosa, cachexia, lipid disorders, and diabetes.

[0077] The 46584 protein, fragments thereof, and derivatives and other variants of the sequence in SEQ ID NO:2 thereof are collectively referred to as "polypeptides or proteins of the invention" or "46584 polypeptides or proteins". Nucleic acid molecules encoding such polypeptides or proteins are collectively referred to as "nucleic acids of the invention" or "46584 nucleic acids."

[0078] As used herein, the term "nucleic acid molecule" includes DNA molecules (e.g., a cDNA or genomic DNA) and RNA molecules (e.g., an mRNA) and analogs of the DNA or RNA generated, e.g., by the use of nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

[0079] The term "isolated or purified nucleic acid molecule" includes nucleic acid molecules which are separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and/or 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of 5' and/or 3' nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0080] As used herein, the term "hybridizes under low stringency, medium stringency, high stringency, or very high stringency conditions" describes conditions for hybridization and washing. Guidance for performing hybridization reactions can be found in *Current Protocols in Molecular Biology* (1989) John Wiley & Sons, N.Y., 6.3.1-6.3.6, which is incorporated by reference. Aqueous and nonaqueous methods are described in that reference and either can be used. Specific hybridization conditions referred to herein are as follows: 1) low stringency hybridization conditions in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by two washes in 0.2X SSC, 0.1% SDS at least at 50°C (the temperature of the washes can be increased to 55°C for low stringency conditions); 2) medium stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C; 3) high stringency hybridization conditions in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C; and preferably 4) very high stringency hybridization conditions are 0.5M sodium phosphate, 7% SDS at 65°C, followed by one or more washes at 0.2X SSC, 1% SDS at 65°C. Very high stringency conditions (4) are the preferred

conditions and the ones that should be used unless otherwise specified

[0081] As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

[0082] As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which include an open reading frame encoding a 46584 protein, preferably a mammalian 46584 protein, and can further include non-coding regulatory sequences, and introns.

[0083] An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means preparation of 46584 protein having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non-46584 protein (also referred to herein as a "contaminating protein"), or of chemical precursors or non-46584 chemicals. When the 46584 protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation. The invention includes isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

[0084] A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence of 46584 (e.g., the sequence of SEQ ID NO:1 or 3) without abolishing or more preferably, without substantially altering a biological activity, whereas an "essential" amino acid residue results in such a change. For example, amino acid residues that are conserved among the polypeptides of the present invention, e.g., those present in the sugar transporter domain, are predicted to be particularly unamenable to alteration.

[0085] A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a 46584 protein is preferably replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a 46584 coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for 46584 biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:1 or SEQ ID NO:3, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

[0086] As used herein, a "biologically active portion" of a 46584 protein includes a fragment of a 46584 protein which participates in an interaction between a 46584 molecule and a non-46584 molecule. Biologically active portions of a 46584 protein include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of the 46584 protein, e.g., the amino acid sequence shown in SEQ ID NO:2, which include fewer amino acids than the full length 46584 protein, and exhibit at least one activity of a 46584 protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the 46584 protein, e.g., (1) the ability to reside within a membrane, e.g., a cell membrane (e.g., a nerve cell membrane, fat cell membrane, muscle cell membrane, and/or blood cell membrane, such as an erythrocyte membrane) or organelle (e.g., mitochondrion) membrane; (2) the ability to interact with a substrate or target molecule, e.g., a monosaccharide, (e.g., glucose, fructose, galactose or myo-inositol) or an organic ion (e.g., an organic anion, an organic cation, a prostaglandin, a steroidal compound, a bile acid, a drug, a neurotransmitter, a sulfated lipophilic metabolite, a glucuronidated lipophilic metabolite, a polyamine, a carnitine, or a choline) across a membrane; (3) the ability to transport the substrate or target molecule across the membrane; (4) the ability to transport a second substrate or target molecule, e.g., an ion, (e.g., a bicarbonate ion or a dicarboxylate ion), across a membrane; (5) the ability to interact with and/or modulate a second non-transporter protein; (6) the ability to modulate cellular signaling and/or gene transcription (e.g., either directly or indirectly); (7) the ability to protect cells and/or tissues from organic ions; (8) the ability to protect cells and/or tissues from organic anions; (9) the ability to modulate hormonal responses, e.g., insulin and/or glucagon secretion; (10) the ability to modulate metabolism; (11) the ability to modulate excretion; and (12) the ability to modulate sugar homeostasis in a cell.

[0087] A biologically active portion of a 46584 protein can be a polypeptide which is, for example, 10, 25, 50, 100, 200 or more amino acids in length. Biologically active portions of a 46584 protein can be used as targets for developing agents which modulate a 46584 mediated activity, e.g., (1) the ability to reside within a membrane, e.g., a cell membrane (e.g., a nerve cell membrane, fat cell membrane, muscle cell membrane, and/or blood cell membrane, such as an erythrocyte membrane) or organelle (e.g., mitochondrion) membrane; (2) the ability to interact with a substrate or target molecule, e.g., a monosaccharide, (e.g., glucose, fructose, galactose or myo-inositol) or an organic ion (e.g., an organic anion, an organic cation, a prostaglandin, a steroidal compound, a bile acid, a drug, a neurotransmitter, a sulfated lipophilic metabolite, a glucuronidated lipophilic metabolite, a polyamine, a carnitine, or a choline) across a membrane;

(3) the ability to transport the substrate or target molecule across the membrane; (4) the ability to transport a second substrate or target molecule, *e.g.*, an ion, (*e.g.*, a bicarbonate ion or a dicarboxylate ion), across a membrane; (5) the ability to interact with and/or modulate a second non-transporter protein; (6) the ability to modulate cellular signaling and/or gene transcription (*e.g.*, either directly or indirectly); (7) the ability to protect cells and/or tissues from organic ions; (8) the ability to protect cells and/or tissues from organic anions; (9) the ability to modulate hormonal responses, *e.g.*, insulin and/or glucagon secretion; (10) the ability to modulate metabolism; (11) the ability to modulate excretion; and (12) the ability to modulate sugar homeostasis in a cell.

[0088] Calculations of homology or sequence identity (the terms "homology" and "identity" are used interchangeably herein) between sequences are performed as follows:

[0089] To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, 90%, 100% of the length of the reference sequence (*e.g.*, when aligning a second sequence to the 46584 amino acid sequence of SEQ ID NO:2 having 727 amino acid residues, at least 218, preferably at least 290, more preferably at least 364, even more preferably at least 436, and even more preferably at least 509, 582, or 654 amino acid residues are aligned). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

[0090] The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (1970) *J. Mol. Biol.* 48:444-453 algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. A particularly preferred set of parameters (and the one that should be used if the practitioner is uncertain about what parameters should be applied to determine if a molecule is within a sequence identity or homology limitation of the invention) are a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frameshift gap penalty of 5.

[0091] The percent identity between two amino acid or nucleotide sequences can be determined using the algorithm of Meyers and Miller ((1989) CABIOS, 4:11-17) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

[0092] The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to 46584 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to 46584 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25:3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

[0093] Particular 46584 polypeptides of the present invention have an amino acid sequence substantially identical to the amino acid sequence of SEQ ID NO:2. In the context of an amino acid sequence, the term "substantially identical" is used herein to refer to a first amino acid that contains a sufficient or minimum number of amino acid residues that are i) identical to, or ii) conservative substitutions of aligned amino acid residues in a second amino acid sequence such that the first and second amino acid sequences can have a common structural domain and/or common functional activity. For example, amino acid sequences that contain a common structural domain having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:2 are termed substantially identical.

[0094] In the context of nucleotide sequence, the term "substantially identical" is used herein to refer to a first nucleic acid sequence that contains a sufficient or minimum number of nucleotides that are identical to aligned nucleotides in

a second nucleic acid sequence such that the first and second nucleotide sequences encode a polypeptide having common functional activity, or encode a common structural polypeptide domain or a common functional polypeptide activity. For example, nucleotide sequences having at least about 60%, or 65% identity, likely 75% identity, more likely 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to SEQ ID NO:1 or 3 are termed substantially identical.

[0095] "Misexpression or aberrant expression", as used herein, refers to a non-wild type pattern of gene expression, at the RNA or protein level. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-translational modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

[0096] "Subject", as used herein, can refer to a mammal, e.g., a human, or to an experimental or animal or disease model. The subject can also be a non-human animal, e.g., a horse, cow, goat, or other domestic animal.

[0097] A "purified preparation of cells", as used herein, refers to, in the case of plant or animal cells, an *in vitro* preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

[0098] Various aspects of the invention are described in further detail below.

Isolated Nucleic Acid Molecules

[0099] In one aspect, the invention provides, an isolated or purified, nucleic acid molecule that encodes a 46584 polypeptide described herein, e.g., a full length 46584 protein or a fragment thereof, e.g., a biologically active portion of 46584 protein. Also included is a nucleic acid fragment suitable for use as a hybridization probe, which can be used, e.g., to identify a nucleic acid molecule encoding a polypeptide of the invention, 46584 mRNA, and fragments suitable for use as primers, e.g., PCR primers for the amplification or mutation of nucleic acid molecules.

[0100] In one embodiment, an isolated nucleic acid molecule of the invention includes the nucleotide sequence shown in SEQ ID NO:1, or a portion of any of this nucleotide sequence. In one embodiment, the nucleic acid molecule includes sequences encoding the human 46584 protein (i.e., "the coding region" of SEQ ID NO:1, as shown in SEQ ID NO:3), as well as 5' untranslated sequences (nucleotides 1 to 194 of SEQ ID NO:1) and 3' untranslated sequences (nucleotides 2379 to 6911 of SEQ ID NO:1). Alternatively, the nucleic acid molecule can include only the coding region of SEQ ID NO:1 (e.g., SEQ ID NO:3) and, e.g., no flanking sequences which normally accompany the subject sequence. In another embodiment, the nucleic acid molecule encodes a sequence corresponding to a fragment of the protein from about amino acid 138 to 727 of SEQ ID NO:2.

[0101] In another embodiment, an isolated nucleic acid molecule of the invention includes a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3, or a portion of any of these nucleotide sequences. In other embodiments, the nucleic acid molecule of the invention is sufficiently complementary to the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3 such that it can hybridize to the nucleotide sequence shown in SEQ ID NO:1 or 3, thereby forming a stable duplex.

[0102] In one embodiment, an isolated nucleic acid molecule of the present invention includes a nucleotide sequence which is at least about 90%, 96%, 97%, 98%, 99%, or more homologous to the entire length of the nucleotide sequence shown in SEQ ID NO: 1 or SEQ ID NO:3, or a portion, preferably of the same length, of any of these nucleotide sequences.

46584 Nucleic Acid Fragments

[0103] A nucleic acid molecule of the invention can include only a portion of the nucleic acid sequence of SEQ ID NO:1 or 3. For example, such a nucleic acid molecule can include a fragment which can be used as a probe or primer or a fragment encoding a portion of a 46584 protein, e.g., an immunogenic or biologically active portion of a 46584 protein. A fragment can comprise those nucleotides of SEQ ID NO:1, which encode a sugar transporter domain of human 46584. The nucleotide sequence determined from the cloning of the 46584 gene allows for the generation of probes and primers designed for use in identifying and/or cloning other 46584 family members, or fragments thereof, as well as 46584 homologs, or fragments thereof, from other species.

[0104] In another embodiment, a nucleic acid includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5' or 3' noncoding region. Other embodiments include a fragment which

includes a nucleotide sequence encoding an amino acid fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof, particularly fragments thereof which are at least 100 amino acids in length. Fragments also include nucleic acid sequences corresponding to specific amino acid sequences described above or fragments thereof. Nucleic acid fragments should not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

[0105] A nucleic acid fragment can include a sequence corresponding to a domain, region, or functional site described herein. A nucleic acid fragment can also include one or more domain, region, or functional site described herein. Thus, for example, a 46584 nucleic acid fragment can include a sequence corresponding to a sugar transporter domain, as described herein.

[0106] 46584 probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7, 12 or 15, preferably about 20 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 consecutive nucleotides of a sense or antisense sequence of SEQ ID NO:1 or SEQ ID NO:3, or of a naturally occurring allelic variant or mutant of SEQ ID NO:1 or SEQ ID NO:3.

[0107] In a preferred embodiment the nucleic acid is a probe which is at least 5 or 10, and less than 200, more preferably less than 100, or less than 50, base pairs in length. It should be identical, or differ by 1, or less than in 5 or 10 bases, from a sequence disclosed herein.

If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[0108] A probe or primer can be derived from the sense or anti-sense strand of a nucleic acid which encodes a sugar transporter domain from about amino acid 138 to 727 of SEQ ID NO:2.

[0109] In another embodiment a set of primers is provided, e.g., primers suitable for use in a PCR, which can be used to amplify a selected region of a 46584 sequence, e.g., a domain, region, site or other sequence described herein. The primers should be at least 5, 10, or 50 base pairs in length and less than 100, or less than 200, base pairs in length. The primers should be identical, or differ by one base from a sequence disclosed herein or from a naturally occurring variant. For example, primers suitable for amplifying all or a portion of any of the following regions are provided: a sugar transporter domain from about amino acid 138 to 727 of SEQ ID NO:2.

[0110] A nucleic acid fragment can encode an epitope bearing region of a polypeptide described herein.

[0111] A nucleic acid fragment encoding a "biologically active portion of a 46584 polypeptide" can be prepared by isolating a portion of the nucleotide sequence of SEQ ID NO:1 or 3, which encodes a polypeptide having a 46584 biological activity (e.g., the biological activities of the 46584 proteins are described herein), expressing the encoded portion of the 46584 protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the 46584 protein. For example, a nucleic acid fragment encoding a biologically active portion of 46584 includes a sugar transporter domain, e.g., amino acid residues about 138 to 727 of SEQ ID NO:2. A nucleic acid fragment encoding a biologically active portion of a 46584 polypeptide, can comprise a nucleotide sequence which is greater than 300 or more nucleotides in length.

[0112] In preferred embodiments, a nucleic acid includes a nucleotide sequence which is about 300, 500, 700, 900, 1000, 1200, 1400, 1600, 1800, 2000 or more nucleotides in length and hybridizes under stringent hybridization conditions to a nucleic acid molecule of SEQ ID NO:1 or SEQ ID NO:3.

46584 Nucleic Acid Variants

[0113] The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO:3. Such differences can be due to degeneracy of the genetic code (and result in a nucleic acid which encodes the same 46584 proteins as those encoded by the nucleotide sequence disclosed herein. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence which differs, by at least 1, but less than 5, 10, 20, 50, or 100 amino acid residues that shown in SEQ ID NO:2. If alignment is needed for this comparison the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[0114] Nucleic acids of the inventor can be chosen for having codons, which are preferred, or non-preferred, for a particular expression system. E.g., the nucleic acid can be one in which at least one codon, at preferably at least 10%, or 20% of the codons has been altered such that the sequence is optimized for expression in *E. coli*, yeast, human, insect, or CHO cells.

[0115] Nucleic acid variants can be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism) or can be non naturally occurring. Non-naturally occurring variants can be made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. The variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions

(as compared in the encoded product).

[0116] In a preferred embodiment, the nucleic acid differs from that of SEQ ID NO: 1 or 3, *e.g.*, as follows: by at least one but less than 10, 20, 30, or 40 nucleotides; at least one but less than 1%, 5%, 10% or 20% of the nucleotides in the subject nucleic acid. If necessary for this analysis the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

[0117] Orthologs, homologs, and allelic variants can be identified using methods known in the art. These variants comprise a nucleotide sequence encoding a polypeptide that is 50%, at least about 55%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more identical to the nucleotide sequence shown in SEQ ID NO:2 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO 2 or a fragment of the sequence. Nucleic acid molecules corresponding to orthologs, homologs, and allelic variants of the 46584 cDNAs of the invention can further be isolated by mapping to the same chromosome or locus as the 46584 gene.

[0118] Preferred variants include those that are correlated with at least one of the following activities: (1) the ability to reside within a membrane, *e.g.*, a cell membrane (*e.g.*, a nerve cell membrane, fat cell membrane, muscle cell membrane, and/or blood cell membrane, such as an erythrocyte membrane) or organelle (*e.g.*, mitochondrion) membrane; (2) the ability to interact with a substrate or target molecule, *e.g.*, a monosaccharide, (*e.g.*, glucose, fructose, galactose or myo-inositol) or an organic ion (*e.g.*, an organic anion, an organic cation, a prostaglandin, a steroidal compound, a bile acid, a drug, a neurotransmitter, a sulfated lipophilic metabolite, a glucuronidated lipophilic metabolite, a polyamine, a carnitine, or a choline) across a membrane; (3) the ability to transport the substrate or target molecule across the membrane; (4) the ability to transport a second substrate or target molecule, *e.g.*, an ion, (*e.g.*, a bicarbonate ion or a dicarboxylate ion), across a membrane; (5) the ability to interact with and/or modulate a second non-transporter protein; (6) the ability to modulate cellular signaling and/or gene transcription (*e.g.*, either directly or indirectly); (7) the ability to protect cells and/or tissues from organic ions; (8) the ability to protect cells and/or tissues from organic anions; (9) the ability to modulate hormonal responses, *e.g.*, insulin and/or glucagon secretion; (10) the ability to modulate metabolism; (11) the ability to modulate excretion; and (12) the ability to modulate sugar homeostasis in a cell.

[0119] Allelic variants of 46584, *e.g.*, human 46584, include both functional and non-functional proteins. Functional allelic variants are naturally occurring amino acid sequence variants of the 46584 protein within a population that maintain at least one of the following activities: (1) the ability to reside within a membrane, *e.g.*, a cell membrane (*e.g.*, a nerve cell membrane, fat cell membrane, muscle cell membrane, and/or blood cell membrane, such as an erythrocyte membrane) or organelle (*e.g.*, mitochondrion) membrane; (2) the ability to interact with a substrate or target molecule, *e.g.*, a monosaccharide, (*e.g.*, glucose, fructose, galactose or myo-inositol) or an organic ion (*e.g.*, an organic anion, an organic cation, a prostaglandin, a steroidal compound, a bile acid, a drug, a neurotransmitter, a sulfated lipophilic metabolite, a glucuronidated lipophilic metabolite, a polyamine, a carnitine, or a choline) across a membrane; (3) the ability to transport the substrate or target molecule across the membrane; (4) the ability to transport a second substrate or target molecule, *e.g.*, an ion, (*e.g.*, a bicarbonate ion or a dicarboxylate ion), across a membrane; (5) the ability to interact with and/or modulate a second non-transporter protein; (6) the ability to modulate cellular signaling and/or gene transcription (*e.g.*, either directly or indirectly); (7) the ability to protect cells and/or tissues from organic ions; (8) the ability to protect cells and/or tissues from organic anions; (9) the ability to modulate hormonal responses, *e.g.*, insulin and/or glucagon secretion; (10) the ability to modulate metabolism; (11) the ability to modulate excretion; and (12) the ability to modulate sugar homeostasis in a cell.

[0120] Functional allelic variants will typically contain only conservative substitution of one or more amino acids of SEQ ID NO:2, or substitution, deletion or insertion of non-critical residues in non-critical regions of the protein. Non-functional allelic variants are naturally-occurring amino acid sequence variants of the 46584, *e.g.*, human 46584, protein within a population that do not have any of the following activities: (1) the ability to reside within a membrane, *e.g.*, a cell membrane (*e.g.*, a nerve cell membrane, fat cell membrane, muscle cell membrane, and/or blood cell membrane, such as an erythrocyte membrane) or organelle (*e.g.*, mitochondrion) membrane; (2) the ability to interact with a substrate or target molecule, *e.g.*, a monosaccharide, (*e.g.*, glucose, fructose, galactose or myo-inositol); (3) the ability to transport the substrate or target molecule across the membrane; (4) the ability to interact with and/or modulate a second non-transporter protein; (5) the ability to modulate sugar homeostasis in a cell; (6) the ability to modulate insulin and/or glucagon secretion; and (7) the ability to modulate metabolism.

[0121] Non-functional allelic variants will typically contain a non-conservative substitution, a deletion, or insertion, or premature truncation of the amino acid sequence of SEQ ID NO:2, or a substitution, insertion, or deletion in critical residues or critical regions of the protein.

[0122] Moreover, nucleic acid molecules encoding other 46584 family members and, thus, which have a nucleotide sequence which differs from the 46584 sequences of SEQ ID NO:1 or SEQ ID NO:3 are intended to be within the scope of the invention.

Antisense Nucleic Acid Molecules, Ribozymes and Modified 46584 Nucleic Acid Molecules

[0123] In another aspect, the invention features, an isolated nucleic acid molecule which is antisense to 46584. An "antisense" nucleic acid can include a nucleotide sequence which is complementary to a "sense" nucleic acid encoding a protein, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. The antisense nucleic acid can be complementary to an entire 46584 coding strand, or to only a portion thereof (e.g., the coding region of human 46584 corresponding to SEQ ID NO:3). In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding 46584 (e.g., the 5' and 3' untranslated regions).

[0124] An antisense nucleic acid can be designed such that it is complementary to the entire coding region of 46584 mRNA, but more preferably is an oligonucleotide which is antisense to only a portion of the coding or noncoding region of 46584 mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of 46584 mRNA, e.g., between the -10 and +10 regions of the target gene nucleotide sequence of interest. An antisense oligonucleotide can be, for example, about 7, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, or more nucleotides in length.

[0125] An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. The antisense nucleic acid also can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

[0126] The antisense nucleic acid molecules of the invention are typically administered to a subject (e.g., by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a 46584 protein to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, antisense molecules can be modified such that they specifically or selectively bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

[0127] In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue *et al.* (1987) *FEBS Lett.* 215:327-330).

[0128] In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. A ribozyme having specificity for a 46584-encoding nucleic acid can include one or more sequences complementary to the nucleotide sequence of a 46584 cDNA disclosed herein (i.e., SEQ ID NO:1 or SEQ ID NO:3), and a sequence having known catalytic sequence responsible for mRNA cleavage (see U.S. Pat. No. 5,093,246 or Haselhoff and Gerlach (1988) *Nature* 334:585-591). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a 46584-encoding mRNA. See, e.g., Cech *et al.* U.S. Patent No. 4,987,071; and Cech *et al.* U.S. Patent No. 5,116,742. Alternatively, 46584 mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, e.g., Bartel and Szostak (1993) *Science* 261:1411-1418.

[0129] 46584 gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the 46584 (e.g., the 46584 promoter and/or enhancers) to form triple helical structures that prevent transcription of the 46584 gene in target cells. See generally, Helene (1991) *Anticancer Drug Des.* 6:569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14:807-15. The potential sequences that can be targeted for triple helix formation can be increased by creating a so-called "switchback" nucleic acid molecule. Switchback molecules are synthesized in an alternating 5'-3', 3'-5' manner, such that they base pair with first one strand of a duplex and then the other, eliminating the necessity for a sizeable stretch of either purines or pyrimidines to be present on one strand of a duplex.

[0130] The invention also provides detectably labeled oligonucleotide primer and probe molecules. Typically, such labels are chemiluminescent, fluorescent, radioactive, or colorimetric.

[0131] A 46584 nucleic acid molecule can be modified at the base moiety, sugar moiety or phosphate backbone to

improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acid molecules can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorganic & Medicinal Chemistry* 4: 5-23). As used herein, the terms "peptide nucleic acid" or "PNA" refers to a nucleic acid mimic, e.g., a DNA mimic, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of a PNA can allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) *supra*; Perry-O'Keefe *et al.* (1996) *Proc. Natl. Acad. Sci.* 93: 14670-675.

[0132] PNAs of 46584 nucleic acid molecules can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, for example, inducing transcription or translation arrest or inhibiting replication. PNAs of 46584 nucleic acid molecules can also be used in the analysis of single base pair mutations in a gene, (e.g., by PNA-directed PCR clamping); as 'artificial restriction enzymes' when used in combination with other enzymes, (e.g., S1 nucleases (Hyrup *et al.* (1996) *supra*)); or as probes or primers for DNA sequencing or hybridization (Hyrup *et al.* (1996) *supra*; Perry-O'Keefe *supra*).

[0133] In other embodiments, the oligonucleotide can include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol *et al.* (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (see, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide can be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

[0134] The invention also includes molecular beacon oligonucleotide primer and probe molecules having at least one region which is complementary to a 46584 nucleic acid of the invention, two complementary regions one having a fluorophore and one a quencher such that the molecular beacon is useful for quantitating the presence of the 46584 nucleic acid of the invention in a sample. Molecular beacon nucleic acids are described, for example, in Lizardi *et al.*, U.S. Patent No. 5,854,033; Nazarenko *et al.*, U.S. Patent No. 5,866,336, and Livak *et al.*, U.S. Patent 5,876,930.

Isolated 46584 Polypeptides

[0135] In another aspect, the invention features, an isolated 46584 protein, or fragment, e.g., a biologically active portion, for use as immunogens or antigens to raise or test (or more generally to bind) anti-46584 antibodies. 46584 protein can be isolated from cells or tissue sources using standard protein purification techniques. 46584 protein or fragments thereof can be produced by recombinant DNA techniques or synthesized chemically.

[0136] Polypeptides of the invention include those which arise as a result of the existence of multiple genes, alternative transcription events, alternative RNA splicing events, and alternative translational and post-translational events. The polypeptide can be expressed in systems, e.g., cultured cells, which result in substantially the same post-translational modifications present when the polypeptide is expressed in a native cell, or in systems which result in the alteration or omission of post-translational modifications, e.g., glycosylation or cleavage, present in a native cell.

[0137] In a preferred embodiment, a 46584 polypeptide has one or more of the following characteristics:

it has the ability to perform one or more of the following activities: (1) the ability to reside within a membrane, e.g., a cell membrane (e.g., a nerve cell membrane, fat cell membrane, muscle cell membrane, and/or blood cell membrane, such as an erythrocyte membrane) or organelle (e.g., mitochondrion) membrane; (2) the ability to interact with a substrate or target molecule, e.g., a monosaccharide, (e.g., glucose, fructose, galactose or myo-inositol) or an organic ion (e.g., an organic anion, an organic cation, a prostaglandin, a steroidal compound, a bile acid, a drug, a neurotransmitter, a sulfated lipophilic metabolite, a glucuronidated lipophilic metabolite, a polyamine, a carnitine, or a choline) across a membrane; (3) the ability to transport the substrate or target molecule across the membrane; (4) the ability to transport a second substrate or target molecule, e.g., an ion, (e.g., a bicarbonate ion or a dicarboxylate ion), across a membrane; (5) the ability to interact with and/or modulate a second non-transporter protein; (6) the ability to modulate cellular signaling and/or gene transcription (e.g., either directly or indirectly); (7) the ability to protect cells and/or tissues from organic ions; (8) the ability to protect cells and/or tissues from organic anions; (9) the ability to modulate hormonal responses, e.g., insulin and/or glucagon secretion; (10) the ability to modulate metabolism; (11) the ability to modulate excretion; and (12) the ability to modulate sugar homeostasis in a cell;

it has a molecular weight, e.g., a deduced molecular weight, preferably ignoring any contribution of post translational modifications, amino acid composition or other physical characteristic of a 46584 polypeptide, e.g., a polypeptide of SEQ ID NO:2;

it has an overall sequence similarity of at least 60%, preferably at least 70%, 80%, 90%, or 95%, more preferably at least 97%, 98%, or 99% with a polypeptide of SEQ ID NO:2;

it has a sugar transporter domain which is preferably about 70%, 80%, 90% or 95% identical to amino acid residues about 138 to 727 of SEQ ID NO:2; or

it has a sugar transport protein signature sequence which is preferably about 70%, 80%, 90% or 95%, more preferably about 97%, 98%, or 99% identical to amino acid residues about 250 to 275 of SEQ ID NO:2.

[0138] In another preferred embodiment, the 46584 protein, or fragment thereof, differs from the corresponding sequence in SEQ ID NO:2. In one embodiment it differs by at least one but by less than 15, 10 or 5 amino acid residues. In another it differs from the corresponding sequence in SEQ ID NO:2 by at least one residue but less than 20%, 15%, 10% or 5% of the residues in it differ from the corresponding sequence in SEQ ID NO:2. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) The differences are, preferably, differences or changes at a non-essential residue or a conservative substitution. In a preferred embodiment the differences are not in the sugar transporter domain at about residues 138 to 727 of SEQ ID NO:2. In another embodiment one or more differences are in the sugar transporter domain at about residues 138 to 727 of SEQ ID NO:2.

[0139] Other embodiments include a protein that contains one or more changes in amino acid sequence, e.g., a change in an amino acid residue which is not essential for activity. Such 46584 proteins differ in amino acid sequence from SEQ ID NO:2, yet retain biological activity.

[0140] [00140] In one embodiment, the protein includes an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% or more homologous to SEQ ID NO:2.

[0141] A 46584 protein or fragment is provided which varies from the sequence of SEQ ID NO:2 in regions defined by amino acids about 1 to 137 by at least one but by less than 15, 10 or 5 amino acid residues in the protein or fragment but which does not differ from SEQ ID NO:2 in regions defined by amino acids about 138 to 727. (If this comparison requires alignment the sequences should be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.) In some embodiments the difference is at a non-essential residue or is a conservative substitution, while in others the difference is at an essential residue or is a non-conservative substitution.

[0142] In one embodiment, a biologically active portion of a 46584 protein includes a sugar transporter domain. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native 46584 protein.

[0143] In a preferred embodiment, the 46584 protein has an amino acid sequence shown in SEQ ID NO:2. In other embodiments, the 46584 protein is sufficiently or substantially identical to SEQ ID NO:2. In yet another embodiment, the 46584 protein is sufficiently or substantially identical to SEQ ID NO:2 and retains the functional activity of the protein of SEQ ID NO:2, as described in detail in the subsections above.

46584 Chimeric or Fusion Proteins

[0144] In another aspect, the invention provides 46584 chimeric or fusion proteins. As used herein, a 46584 "chimeric protein" or "fusion protein" includes a 46584 polypeptide linked to a non-46584 polypeptide. A "non-46584 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the 46584 protein, e.g., a protein which is different from the 46584 protein and which is derived from the same or a different organism. The 46584 polypeptide of the fusion protein can correspond to all or a portion e.g., a fragment described herein of a 46584 amino acid sequence. In a preferred embodiment, a 46584 fusion protein includes at least one (or two) biologically active portion of a 46584 protein. The non-46584 polypeptide can be fused to the N-terminus or C-terminus of the 46584 polypeptide.

[0145] The fusion protein can include a moiety which has a high affinity for a ligand. For example, the fusion protein can be a GST-46584 fusion protein in which the 46584 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant 46584. Alternatively, the fusion protein can be a 46584 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of 46584 can be increased through use of a heterologous signal sequence.

[0146] Fusion proteins can include all or a part of a serum protein, e.g., a portion of an immunoglobulin (e.g., IgG, IgA, or IgE), e.g., an Fc region and/or the hinge C1 and C2 sequences of an immunoglobulin or human serum albumin.

[0147] The 46584 fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. The 46584 fusion proteins can be used to affect the bioavailability of a 46584 substrate. 46584 fusion proteins can be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a 46584 protein; (ii) mis-regulation of the 46584 gene; and (iii) aberrant post-translational modification of a 46584 protein.

[0148] Moreover, the 46584-fusion proteins of the invention can be used as immunogens to produce anti-46584 antibodies in a subject, to purify 46584 ligands and in screening assays to identify molecules which inhibit the interaction of 46584 with a 46584 substrate.

[0149] Expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A 46584-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the 46584 protein. Variants of 46584 Proteins

[0150] In another aspect, the invention also features a variant of a 46584 polypeptide, e.g., which functions as an agonist (mimetics) or as an antagonist. Variants of the 46584 proteins can be generated by mutagenesis, e.g., discrete point mutation, the insertion or deletion of sequences or the truncation of a 46584 protein. An agonist of the 46584 proteins can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of a 46584 protein. An antagonist of a 46584 protein can inhibit one or more of the activities of the naturally occurring form of the 46584 protein by, for example, competitively modulating a 46584-mediated activity of a 46584 protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Preferably, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the 46584 protein.

[0151] Variants of a 46584 protein can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of a 46584 protein for agonist or antagonist activity.

[0152] Libraries of fragments e.g., N terminal, C terminal, or internal fragments, of a 46584 protein coding sequence can be used to generate a variegated population of fragments for screening and subsequent selection of variants of a 46584 protein.

[0153] Variants in which a cysteine residues is added or deleted or in which a residue which is glycosylated is added or deleted are particularly preferred.

[0154] Methods for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property are known in the art. Recursive ensemble mutagenesis (REM), a new technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify 46584 variants (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave *et al.* (1993) *Protein Engineering* 6:327-331).

[0155] Cell based assays can be exploited to analyze a variegated 46584 library. For example, a library of expression vectors can be transfected into a cell line, e.g., a cell line, which ordinarily responds to 46584 in a substrate-dependent manner. The transfected cells are then contacted with 46584 and the effect of the expression of the mutant on signaling by the 46584 substrate can be detected, e.g., by measuring the effect of the mutant on (1) the ability to reside within a membrane, e.g., a cell membrane (e.g., a nerve cell membrane, fat cell membrane, muscle cell membrane, and/or blood cell membrane, such as an erythrocyte membrane) or organelle (e.g., mitochondrion) membrane; (2) the ability to interact with a substrate or target molecule, e.g., a monosaccharide, (e.g., glucose, fructose, galactose or myo-inositol) or an organic ion (e.g., an organic anion, an organic cation, a prostaglandin, a steroidal compound, a bile acid, a drug, a neurotransmitter, a sulfated lipophilic metabolite, a glucuronidated lipophilic metabolite, a polyamine, a carnitine, or a choline) across a membrane; (3) the ability to transport the substrate or target molecule across the membrane; (4) the ability to transport a second substrate or target molecule, e.g., an ion, (e.g., a bicarbonate ion or a dicarboxylate ion), across a membrane; (5) the ability to interact with and/or modulate a second non-transporter protein; (6) the ability to modulate cellular signaling and/or gene transcription (e.g., either directly or indirectly); (7) the ability to protect cells and/or tissues from organic ions; (8) the ability to protect cells and/or tissues from organic anions; (9) the ability to modulate hormonal responses, e.g., insulin and/or glucagon secretion; (10) the ability to modulate metabolism; (11) the ability to modulate excretion; and (12) the ability to modulate sugar homeostasis in a cell.

[0156] Plasmid DNA can then be recovered from the cells which score for inhibition, or alternatively, potentiation of signaling by the 46584 substrate, and the individual clones further characterized.

[0157] In another aspect, the invention features a method of making a 46584 polypeptide, e.g., a peptide having a non-wild type activity, e.g., an antagonist, agonist, or super agonist of a naturally occurring 46584 polypeptide, e.g., a naturally occurring 46584 polypeptide. The method includes altering the sequence of a 46584 polypeptide, e.g., altering the sequence, e.g., by substitution or deletion of one or more residues of a non-conserved region, a domain or residue disclosed herein, and testing the altered polypeptide for the desired activity.

[0158] In another aspect, the invention features a method of making a fragment or analog of a 46584 polypeptide a biological activity of a naturally occurring 46584 polypeptide. The method includes altering the sequence, e.g., by substitution or deletion of one or more residues, of a 46584 polypeptide, e.g., altering the sequence of a non-conserved region, or a domain or residue described herein, and testing the altered polypeptide for the desired activity.

Anti-46584 Antibodies

[0159] In another aspect, the invention provides an anti-46584 antibody. The term "antibody" as used herein refers

to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include scFV and dcFV fragments, Fab and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as papain or pepsin, respectively.

[0160] The antibody can be a polyclonal, monoclonal, recombinant, e.g., a chimeric or humanized, fully human, non-human, e.g., murine, or single chain antibody. In a preferred embodiment it has effector function and can fix complement. The antibody can be coupled to a toxin or imaging agent.

[0161] A full-length 46584 protein or, antigenic peptide fragment of 46584 can be used as an immunogen or can be used to identify anti-46584 antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. The antigenic peptide of 46584 should include at least 8 amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of 46584. Preferably, the antigenic peptide includes at least 10 amino acid residues, more preferably at least 15 amino acid residues, even more preferably at least 20 amino acid residues, and most preferably at least 30 amino acid residues.

[0162] Fragments of 46584 which include residues at about about 1 to 151, 213 to 220, 268 to 278, 341 to 435, 604 to 609, 652 to 667, 719 to 727, 177 to 191, 239 to 248, 302 to 320, 455 to 580, 629 to 634, and 693 to 698 of SEQ ID NO:2 can be used to make, e.g., used as immunogens or used to characterize the specificity of an antibody, antibodies against hydrophilic regions of the 46584 protein (for example, shown in Figure 1). Similarly, fragments of 46584 which include residues at about 152 to 176, 192 to 212, 221 to 238, 249 to 267, 279 to 301, 321 to 340, 436 to 454, 581 to 603, 610 to 628, 635 to 651, 668 to 692, and 699 to 718 of SEQ ID NO:2 can be used to make an antibody against a hydrophobic region of the 46584 protein; fragments of 46584 which include residues 177 to 191, 239 to 248, 302 to 320, 455 to 580, 629 to 634, and 693 to 698 of SEQ ID NO:2 can be used to make an antibody against an extracellular region of the 46584 protein; fragments of 46584 which include residues at about 1 to 151, 213 to 220, 268 to 278, 341 to 435, 604 to 609, 652 to 667, and 719 to 727 of SEQ ID NO:2 can be used to make an antibody against an intracellular region of the 46584 protein; a fragment of 46584 which include residues at about 138 to 727 of SEQ ID NO:2 can be used to make an antibody against the sugar transporter domain region of the 46584 protein.

[0163] Antibodies reactive with, or specific or selective for, any of these regions, or other regions or domains described herein are provided.

[0164] Preferred epitopes encompassed by the antigenic peptide are regions of 46584 located on the surface of the protein, e.g., hydrophilic regions, as well as regions with high antigenicity. For example, an Emini surface probability analysis of the human 46584 protein sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface of the 46584 protein and are thus likely to constitute surface residues useful for targeting antibody production.

[0165] In a preferred embodiment the antibody can bind to an extracellular portion of the 46584 protein, e.g., it can bind to a whole cell which expresses the 46584 protein. In another embodiment, the antibody binds an intracellular portion of the 46584 protein.

[0166] In a preferred embodiment the antibody binds an epitope on any domain or region on 46584 proteins described herein.

[0167] Additionally, chimeric, humanized, and completely human antibodies are also within the scope of the invention. Chimeric, humanized, but most preferably, completely human antibodies are desirable for applications which include repeated administration, e.g., therapeutic treatment of human patients, and some diagnostic applications.

[0168] Chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, can be made using standard recombinant DNA techniques. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in Robinson *et al.* International Application No. PCT/US86/02269; Akira, *et al.* European Patent Application 184,187; Taniguchi, European Patent Application 171,496; Morrison *et al.* European Patent Application 173,494; Neuberger *et al.* PCT International Publication No. WO 86/01533; Cabilly *et al.* U.S. Patent No. 4,816,567; Cabilly *et al.* European Patent Application 125,023; Better *et al.* (1988) *Science* 240:1041-1043; Liu *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu *et al.* (1987) *J. Immunol.* 139:3521-3526; Sun *et al.* (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura *et al.* (1987) *Canc. Res.* 47:999-1005; Wood *et al.* (1985) *Nature* 314:446-449; and Shaw *et al.* (1988) *J. Natl. Cancer Inst.* 80:1553-1559; Morrison (1985) *Science* 229:1202-1207; Oi *et al.* (1986) *BioTechniques* 4:214; Winter U.S. Patent 5,225,539; Jones *et al.* (1986) *Nature* 321:552-525; Verhoeyan *et al.* (1988) *Science* 239:1534; and Beidler *et al.* (1988) *J. Immunol.* 141:4053-4060.

[0169] Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced using transgenic mice that are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. See, for example, Lonberg and Huszar (1995) *Int. Rev. Immunol.* 13:65-93; and U.S. Patent Nos. 5,625,126; 5,633,425; 5,569,825; 5,661,016; and 5,545,806. In addition, companies such as Abgenix, Inc. (Fremont, CA) and Medarex, Inc. (Princeton, NJ), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

[0170] Completely human antibodies that recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, e.g., a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. This technology is described by Jespers *et al.* (1994) *Bio/Technology* 12:899-903).

[0171] The anti-46584 antibody can be a single chain antibody. A single-chain antibody (scFV) can be engineered as described in, for example, Colcher *et al.* (1999) *Ann. N Y Acad. Sci.* 880:263-80; and Reiter (1996) *Clin. Cancer Res.* 2:245-52. The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target 46584 protein.

[0172] In a preferred embodiment, the antibody has reduced or no ability to bind an Fc receptor. For example, it is an isotype or subtype, fragment or other mutant, which does not support binding to an Fc receptor, e.g., it has a mutagenized or deleted Fc receptor binding region.

[0173] An antibody (or fragment thereof) may be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, teniposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, puromycin, maytansinoids, e.g., maytansinol (see US Patent No. 5,208,020), CC-1065 (see US Patent Nos. 5,475,092, 5,585,499, 5,846,545) and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (e.g., mechlorethamine, thioepa chlorambucil, CC-1065, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclophosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (e.g., dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (e.g., vincristine, vinblastine, taxol and maytansinoids). Radioactive ions include, but are not limited to iodine, yttrium and praseodymium.

[0174] The conjugates of the invention can be used for modifying a given biological response, the therapeutic moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the therapeutic moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, α -interferon, β -interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophage colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

[0175] Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

[0176] An anti-46584 antibody (e.g., monoclonal antibody) can be used to isolate 46584 by standard techniques, such as affinity chromatography or immunoprecipitation.

Moreover, an anti-46584 antibody can be used to detect 46584 protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-46584 antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labelling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

[0177] In preferred embodiments, an antibody can be made by immunizing with a purified 46584 antigen, or a fragment thereof, e.g., a fragment described herein, a membrane associated antigen, tissues, e.g., crude tissue preparations, whole cells, preferably living cells, lysed cells, or cell fractions, e.g., membrane fractions.

[0178] Antibodies which bind only a native 46584 protein, only denatured or otherwise non-native 46584 protein, or which bind both, are within the invention. Antibodies with linear or conformational epitopes are within the invention. Conformational epitopes sometimes can be identified by identifying antibodies which bind to native but not denatured 46584 protein.

Recombinant Expression Vectors, Host Cells and Genetically Engineered Cells

[0179] In another aspect, the invention includes, vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide described herein. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked and can include a plasmid, cosmid or viral vector. The vector can be capable of autonomous replication or it can integrate into a host DNA. Viral vectors include, e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses.

[0180] A vector can include a 46584 nucleic acid in a form suitable for expression of the nucleic acid in a host cell. Preferably the recombinant expression vector includes one or more regulatory sequences operatively linked to the nucleic acid sequence to be expressed. The term "regulatory sequence" includes promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or polypeptides, including fusion proteins or polypeptides, encoded by nucleic acids as described herein (e.g., 46584 proteins, mutant forms of 46584 proteins, fusion proteins, and the like).

[0181] The recombinant expression vectors of the invention can be designed for expression of 46584 proteins in prokaryotic or eukaryotic cells. For example, polypeptides of the invention can be expressed in *E. coli*, insect cells (e.g., using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[0182] Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

[0183] Purified fusion proteins can be used in 46584 activity assays, (e.g., direct assays or competitive assays described in detail below), or to generate antibodies specific or selective for 46584 proteins. In a preferred embodiment, a fusion protein expressed in a retroviral expression vector of the present invention can be used to infect bone marrow cells which are subsequently transplanted into irradiated recipients. The pathology of the subject recipient is then examined after sufficient time has passed (e.g., six weeks).

[0184] To maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada *et al.*, (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

[0185] The 46584 expression vector can be a yeast expression vector, a vector for expression in insect cells, e.g., a baculovirus expression vector or a vector suitable for expression in mammalian cells.

[0186] When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40.

[0187] In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert *et al.* (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji *et al.* (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund *et al.* (1985) *Science* 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-

regulated promoters are also encompassed, for example, the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the α -fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

[0188] The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. Regulatory sequences (e.g., viral promoters and/or enhancers) operatively linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the constitutive, tissue specific or cell type specific expression of antisense RNA in a variety of cell types. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus. For a discussion of the regulation of gene expression using antisense genes see Weintraub *et al.*, (1986) *Reviews - Trends in Genetics* 1:1.

[0189] Another aspect the invention provides a host cell which includes a nucleic acid molecule described herein, e.g., a 46584 nucleic acid molecule within a recombinant expression vector or a 46584 nucleic acid molecule containing sequences which allow it to homologously recombine into a specific site of the host cell's genome. The terms "host cell" and "recombinant host cell" are used interchangeably herein. Such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications can occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0190] A host cell can be any prokaryotic or eukaryotic cell. For example, a 46584 protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary (CHO) cells or CV-1 origin, SV-40 (COS) cells). Other suitable host cells are known to those skilled in the art.

[0191] Vector DNA can be introduced into host cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride coprecipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation.

[0192] A host cell of the invention can be used to produce (i.e., express) a 46584 protein. Accordingly, the invention further provides methods for producing a 46584 protein using the host cells of the invention. In one embodiment, the method includes culturing the host cell of the invention (into which a recombinant expression vector encoding a 46584 protein has been introduced) in a suitable medium such that a 46584 protein is produced. In another embodiment, the method further includes isolating a 46584 protein from the medium or the host cell.

[0193] In another aspect, the invention features, a cell or purified preparation of cells which include a 46584 transgene, or which otherwise misexpress 46584. The cell preparation can consist of human or non-human cells, e.g., rodent cells, e.g., mouse or rat cells, rabbit cells, or pig cells. In preferred embodiments, the cell or cells include a 46584 transgene, e.g., a heterologous form of a 46584, e.g., a gene derived from humans (in the case of a non-human cell). The 46584 transgene can be misexpressed, e.g., overexpressed or underexpressed. In other preferred embodiments, the cell or cells include a gene which misexpresses an endogenous 46584, e.g., a gene the expression of which is disrupted, e.g., a knockout. Such cells can serve as a model for studying disorders which are related to mutated or misexpressed 46584 alleles or for use in drug screening.

[0194] In another aspect, the invention features, a human cell, e.g., a hematopoietic stem cell, transformed with nucleic acid which encodes a subject 46584 polypeptide.

[0195] Also provided are cells, preferably human cells, e.g., human hematopoietic or fibroblast cells, in which an endogenous 46584 is under the control of a regulatory sequence that does not normally control the expression of the endogenous 46584 gene. The expression characteristics of an endogenous gene within a cell, e.g., a cell line or microorganism, can be modified by inserting a heterologous DNA regulatory element into the genome of the cell such that the inserted regulatory element is operably linked to the endogenous 46584 gene. For example, an endogenous 46584 gene which is "transcriptionally silent," e.g., not normally expressed, or expressed only at very low levels, can be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell. Techniques such as targeted homologous recombinations, can be used to insert the heterologous DNA as described in, e.g., Chappel, US 5,272,071; WO 91/06667, published in May 16, 1991.

Transgenic Animals

[0196] The invention provides non-human transgenic animals. Such animals are useful for studying the function and/or activity of a 46584 protein and for identifying and/or evaluating modulators of 46584 activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene is exogenous DNA or a rearrangement, e.g., a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic animal. A transgene can direct the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal, other transgenes, e.g., a knockout, reduce expression. Thus, a transgenic animal can be one in which an endogenous 46584 gene has been altered by, e.g., by homologous

recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

[0197] Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to a transgene of the invention to direct expression of a 46584 protein to particular cells. A transgenic founder animal can be identified based upon the presence of a 46584 transgene in its genome and/or expression of 46584 mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a 46584 protein can further be bred to other transgenic animals carrying other transgenes.

[0198] 46584 proteins or polypeptides can be expressed in transgenic animals or plants, e.g., a nucleic acid encoding the protein or polypeptide can be introduced into the genome of an animal. In preferred embodiments the nucleic acid is placed under the control of a tissue specific promoter, e.g., a milk or egg specific promoter, and recovered from the milk or eggs produced by the animal. Suitable animals are mice, pigs, cows, goats, and sheep.

[0199] The invention also includes a population of cells from a transgenic animal, as discussed, e.g., below.

Uses

[0200] The nucleic acid molecules, proteins, protein homologs, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenetics); and c) methods of treatment (e.g., therapeutic and prophylactic).

[0201] The isolated nucleic acid molecules of the invention can be used, for example, to express a 46584 protein (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect a 46584 mRNA (e.g., in a biological sample) or a genetic alteration in a 46584 gene, and to modulate 46584 activity, as described further below. The 46584 proteins can be used to treat disorders characterized by insufficient or excessive production of a 46584 substrate or production of 46584 inhibitors. In addition, the 46584 proteins can be used to screen for naturally occurring 46584 substrates, to screen for drugs or compounds which modulate 46584 activity, as well as to treat disorders characterized by insufficient or excessive production of 46584 protein or production of 46584 protein forms which have decreased, aberrant or unwanted activity compared to 46584 wild type protein (e.g., aberrant or deficient sugar transport function or expression). Moreover, the anti-46584 antibodies of the invention can be used to detect and isolate 46584 proteins, regulate the bioavailability of 46584 proteins, and modulate 46584 activity.

[0202] A method of evaluating a compound for the ability to interact with, e.g., bind, a subject 46584 polypeptide is provided. The method includes: contacting the compound with the subject 46584 polypeptide; and evaluating ability of the compound to interact with, e.g., to bind or form a complex with the subject 46584 polypeptide. This method can be performed *in vitro*, e.g., in a cell free system, or *in vivo*, e.g., in a two-hybrid interaction trap assay. This method can be used to identify naturally occurring molecules which interact with subject 46584 polypeptide. It can also be used to find natural or synthetic inhibitors of subject 46584 polypeptide. Screening methods are discussed in more detail below.

Screening Assays:

[0203] The invention provides methods (also referred to herein as "screening assays") for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to 46584 proteins, have a stimulatory or inhibitory effect on, for example, 46584 expression or 46584 activity, or have a stimulatory or inhibitory effect on, for example, the expression or activity of a 46584 substrate. Compounds thus identified can be used to modulate the activity of target gene products (e.g., 46584 genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions.

[0204] In one embodiment, the invention provides assays for screening candidate or test compounds which are substrates of a 46584 protein or polypeptide or a biologically active portion thereof. In another embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of a 46584 protein or polypeptide or a biologically active portion thereof.

[0205] The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive; see, e.g., Zuckermann *et al.* (1994) *J. Med. Chem.* 37:2678-85); spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches

are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.* 12:145).

[0206] Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909-13; Erb *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:11422-426; Zuckermann *et al.* (1994). *J Med. Chem.* 37:2678-85; Cho *et al.* (1993) *Science* 261:1303; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carrell *et al.* (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and in Gallop *et al.* (1994) *J. Med. Chem.* 37:1233-51.

[0207] Libraries of compounds can be presented in solution (e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner, USP 5,223,409), spores (Ladner USP '409), plasmids (Cull *et al.* (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla *et al.* (1990) *Proc. Natl. Acad. Sci.* 87: 6378-6382; Felici (1991) *J. Mol. Biol.* 222:301-310; Ladner *supra.*).

[0208] In one embodiment, an assay is a cell-based assay in which a cell which expresses a protein or biologically active portion thereof is contacted with a test compound, and the ability of the test compound to modulate activity is determined. Determining the ability of the test compound to modulate activity can be accomplished by monitoring, for example, one or more of the following activities: (1) the ability to reside within a membrane; (2) the ability to interact with a substrate or target molecule; (3) the ability to transport a substrate or target molecule, e.g., an ion, e.g., an organic ion (e.g., an organic anion, an organic cation, a prostaglandin, a steroidal compound, a bile acid, a drug, a neurotransmitter, a sulfated lipophilic metabolite, a glucuronidated lipophilic metabolite, a polyamine, a carnitine, or a choline) across a membrane; (4) the ability to transport a second substrate or target molecule, e.g., an ion, (e.g., a bicarbonate ion or a dicarboxylate ion), across a membrane; (5) the ability to interact with and/or modulate a second non-transporter protein; (6) the ability to modulate cellular signaling and/or gene transcription (e.g., either directly or indirectly); (7) the ability to protect cells and/or tissues from organic ions; (8) the ability to protect cells and/or tissues from organic anions; (9) the ability to modulate hormonal responses; (10) the ability to modulate metabolism; and (11) the ability to modulate excretion.

[0209] The cell, for example, can be of mammalian origin, e.g., human.

[0210] The ability of the test compound to modulate binding to a compound, e.g., a substrate, or to bind to a substrate can also be evaluated. This can be accomplished, for example, by coupling the compound, e.g., the substrate, with a radioisotope or enzymatic label such that binding of the compound, e.g., the substrate, to a substrate can be determined by detecting the labeled compound, e.g., substrate, in a complex. Alternatively, a substrate could be coupled with a radioisotope or enzymatic label to monitor the ability of a test compound to modulate binding to a substrate in a complex. For example, compounds (e.g., substrates) can be labeled with ¹²⁵I, ¹⁴C, ³⁵S or ³H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

[0211] The ability of a compound (e.g., a substrate) to interact with a protein or biologically active portion thereof with or without the labeling of any of the interactants can be evaluated. For example, a microphysiometer can be used to detect the interaction of a compound with a protein or biologically active portion thereof without the labeling of either the compound or the protein or biologically active portion thereof. McConnell *et al.* (1992) *Science* 257: 1906-1912. As used herein, a "microphysiometer" (e.g., Cytosensor) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between a compound and a protein or biologically active portion thereof.

[0212] In yet another embodiment, a cell-free assay is provided in which a protein or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the protein or biologically active portion thereof is evaluated. Preferred biologically active portions of the proteins to be used in assays of the present invention include fragments which participate in interactions with non-molecule molecules, e.g., fragments with high surface probability scores.

[0213] Soluble and/or membrane-bound forms of isolated proteins (e.g., proteins or biologically active portions thereof) can be used in the cell-free assays of the invention. When membrane-bound forms of the protein are used, it may be desirable to utilize a solubilizing agent. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecylpoly(ethylene glycol ether)_n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

[0214] Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected.

[0215] The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET)

(see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos, *et al.*, U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule can simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label can be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

[0216] In another embodiment, determining the ability of the 46584 protein to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander and Urbaniczky (1991) *Anal. Chem.* 63:2338-2345 and Szabo *et al.* (1995) *Curr. Opin. Struct. Biol.* 5:699-705). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

[0217] In one embodiment, the target gene product or the test substance is anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

[0218] It may be desirable to immobilize either 46584, an anti-46584 antibody or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a 46584 protein, or interaction of a 46584 protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/46584 fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or 46584 protein, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of 46584 binding or activity determined using standard techniques.

[0219] Other techniques for immobilizing either a 46584 protein or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated 46584 protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

[0220] In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific or selective for the immobilized component (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody).

[0221] In one embodiment, this assay is performed utilizing antibodies reactive with 46584 protein or target molecules but which do not interfere with binding of the 46584 protein to its target molecule. Such antibodies can be derivatized to the wells of the plate, and unbound target or 46584 protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the 46584 protein or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the 46584 protein or target molecule.

[0222] Alternatively, cell free assays can be conducted in a liquid phase. In such an assay, the reaction products are separated from unreacted components, by any of a number of standard techniques, including but not limited to: differential centrifugation (see, for example, Rivas and Minton (1993) *Trends Biochem Sci* 18:284-7); chromatography (gel filtration chromatography, ion-exchange chromatography); electrophoresis (see, e.g., Ausubel *et al.*, eds. (1999) *Current Protocols in Molecular Biology*, J. Wiley, New York.); and immunoprecipitation (see, for example, Ausubel *et al.*,

eds. (1999) *Current Protocols in Molecular Biology*, J. Wiley, New York). Such resins and chromatographic techniques are known to one skilled in the art (see, e.g., Heegaard (1998) *J Mol Recognit* 11:141-8; Hage and Tweed (1997) *J Chromatogr B Biomed Sci Appl.* 699:499-525) Further, fluorescence energy transfer can also be conveniently utilized, as described herein, to detect binding without further purification of the complex from solution.

[0223] In a preferred embodiment, the assay includes contacting the 46584 protein or biologically active portion thereof with a known compound which binds 46584 to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a 46584 protein, wherein determining the ability of the test compound to interact with a 46584 protein includes determining the ability of the test compound to preferentially bind to 46584 or biologically active portion thereof, or to modulate the activity of a target molecule, as compared to the known compound.

[0224] The target gene products of the invention can, *in vivo*, interact with one or more cellular or extracellular macromolecules, such as proteins. For the purposes of this discussion, such cellular and extracellular macromolecules are referred to herein as "binding partners." Compounds that disrupt such interactions can be useful in regulating the activity of the target gene product. Such compounds can include, but are not limited to molecules such as antibodies, peptides, and small molecules. The preferred target genes/products for use in this embodiment are the 46584 genes herein identified. In an alternative embodiment, the invention provides methods for determining the ability of the test compound to modulate the activity of a 46584 protein through modulation of the activity of a downstream effector of a 46584 target molecule. For example, the activity of the effector molecule on an appropriate target can be determined, or the binding of the effector to an appropriate target can be determined, as previously described.

[0225] To identify compounds that interfere with the interaction between the target gene product and its cellular or extracellular binding partner(s), a reaction mixture containing the target gene product and the binding partner is prepared, under conditions and for a time sufficient, to allow the two products to form complex. In order to test an inhibitory agent, the reaction mixture is provided in the presence and absence of the test compound. The test compound can be initially included in the reaction mixture, or can be added at a time subsequent to the addition of the target gene and its cellular or extracellular binding partner. Control reaction mixtures are incubated without the test compound or with a placebo. The formation of any complexes between the target gene product and the cellular or extracellular binding partner is then detected. The formation of a complex in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the interaction of the target gene product and the interactive binding partner. Additionally, complex formation within reaction mixtures containing the test compound and normal target gene product can also be compared to complex formation within reaction mixtures containing the test compound and mutant target gene product. This comparison can be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal target gene products.

[0226] These assays can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the target gene product or the binding partner onto a solid phase, and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the target gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are briefly described below.

[0227] In a heterogeneous assay system, either the target gene product or the interactive cellular or extracellular binding partner, is anchored onto a solid surface (e.g., a microtiter plate), while the non-anchored species is labeled, either directly or indirectly. The anchored species can be immobilized by non-covalent or covalent attachments. Alternatively, an immobilized antibody specific or selective for the species to be anchored can be used to anchor the species to the solid surface.

[0228] In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific or selective for the initially non-immobilized species (the antibody, in turn, can be directly labeled or indirectly labeled with, e.g., a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

[0229] Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific or selective for one of the binding components to anchor any complexes formed in solution, and a

labeled antibody specific or selective for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

[0230] In an alternate embodiment of the invention, a homogeneous assay can be used. For example, a preformed complex of the target gene product and the interactive cellular or extracellular binding partner product is prepared in that either the target gene products or their binding partners are labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Patent No. 4,109,496 that utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt target gene product-binding partner interaction can be identified.

[0231] In yet another aspect, the 46584 proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with 46584 ("46584-binding proteins" or "46584-bp") and are involved in 46584 activity. Such 46584-bps can be activators or inhibitors of signals by the 46584 proteins or 46584 targets as, for example, downstream elements of a 46584-mediated signaling pathway.

[0232] The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a 46584 protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. (Alternatively the: 46584 protein can be the fused to the activator domain.) If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a 46584-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., lacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the 46584 protein.

[0233] In another embodiment, modulators of 46584 expression are identified. For example, a cell or cell free mixture is contacted with a candidate compound and the expression of 46584 mRNA or protein evaluated relative to the level of expression of 46584 mRNA or protein in the absence of the candidate compound. When expression of 46584 mRNA or protein is greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of 46584 mRNA or protein expression. Alternatively, when expression of 46584 mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of 46584 mRNA or protein expression. The level of 46584 mRNA or protein expression can be determined by methods described herein for detecting 46584 mRNA or protein.

[0234] In another aspect, the invention pertains to a combination of two or more of the assays described herein. For example, a modulating agent can be identified using a cell-based or a cell free assay, and the ability of the agent to modulate the activity of a 46584 protein can be confirmed *in vivo*, e.g., in an animal such as an animal model for aberrant or deficient sugar transport function or expression.

[0235] This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein (e.g., a 46584 modulating agent, an antisense 46584 nucleic acid molecule, a 46584-specific antibody, or a 46584-binding partner) in an appropriate animal model to determine the efficacy, toxicity, side effects, or mechanism of action, of treatment with such an agent. Furthermore, novel agents identified by the above-described screening assays can be used for treatments as described herein.

Detection Assays

[0236] Portions or fragments of the nucleic acid sequences identified herein can be used as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome e.g., to locate gene regions associated with genetic disease or to associate 46584 with a disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

Chromosome Mapping

[0237] The 46584 nucleotide sequences or portions thereof can be used to map the location of the 46584 genes on a chromosome. This process is called chromosome mapping. Chromosome mapping is useful in correlating the 46584

sequences with genes associated with disease.

[0238] Briefly, 46584 genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the 46584 nucleotide sequences. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the 46584 sequences will yield an amplified fragment.

[0239] A panel of somatic cell hybrids in which each cell line contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, can allow easy mapping of individual genes to specific human chromosomes. (D'Eustachio *et al.* (1983) *Science* 220:919-924).

[0240] Other mapping strategies *e.g.*, in situ hybridization (described in Fan *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries can be used to map 46584 to a chromosomal location.

[0241] Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases will suffice to get good results at a reasonable amount of time. For a review of this technique, see Verma *et al.* (1988) *Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York).

[0242] Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

[0243] Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in McKusick, *Mendelian Inheritance in Man*, available on-line through Johns Hopkins University Welch Medical Library). The relationship between a gene and a disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, for example, Egeland *et al.* (1987) *Nature*, 325: 783-787.

[0244] Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the 46584 gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

[0245] 46584 sequences can be used to identify individuals from biological samples using, *e.g.*, restriction fragment length polymorphism (RFLP). In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, the fragments separated, *e.g.*, in a Southern blot, and probed to yield bands for identification. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

[0246] Furthermore, the sequences of the present invention can also be used to determine the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the 46584 nucleotide sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it. Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences.

[0247] Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO: 1 can provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

[0248] If a panel of reagents from 46584 nucleotide sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely

small tissue samples.

Use of Partial 46584 Sequences in Forensic Biology

[0249] DNA-based identification techniques can also be used in forensic biology. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

[0250] The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions of SEQ ID NO:1 (e.g., fragments derived from the noncoding regions of SEQ ID NO:1 having a length of at least 20 bases, preferably at least 30 bases) are particularly appropriate for this use.

[0251] The 46584 nucleotide sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such 46584 probes can be used to identify tissue by species and/or by organ type.

[0252] In a similar fashion, these reagents, e.g., 46584 primers or probes can be used to screen tissue culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Predictive Medicine

[0253] The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual.

[0254] Generally, the invention provides, a method of determining if a subject is at risk for a disorder related to a lesion in or the misexpression of a gene which encodes 46584.

[0255] Such disorders include, e.g., a disorder associated with the misexpression of 46584 gene.

[0256] The method includes one or more of the following:

detecting, in a tissue of the subject, the presence or absence of a mutation which affects the expression of the 46584 gene, or detecting the presence or absence of a mutation in a region which controls the expression of the gene, e.g., a mutation in the 5' control region;

detecting, in a tissue of the subject, the presence or absence of a mutation which alters the structure of the 46584 gene;

detecting, in a tissue of the subject, the misexpression of the 46584 gene, at the mRNA level, e.g., detecting a non-wild type level of an mRNA; or

detecting, in a tissue of the subject, the misexpression of the gene, at the protein level, e.g., detecting a non-wild type level of a 46584 polypeptide.

[0257] In preferred embodiments the method includes: ascertaining the existence of at least one of: a deletion of one or more nucleotides from the 46584 gene; an insertion of one or more nucleotides into the gene, a point mutation, e.g., a substitution of one or more nucleotides of the gene, a gross chromosomal rearrangement of the gene, e.g., a translocation, inversion, or deletion.

[0258] For example, detecting the genetic lesion can include: (i) providing a probe/primer including an oligonucleotide containing a region of nucleotide sequence which hybridizes to a sense or antisense sequence from SEQ ID NO:1, or naturally occurring mutants thereof or 5' or 3' flanking sequences naturally associated with the 46584 gene; (ii) exposing the probe/primer to nucleic acid of the tissue; and detecting, by hybridization, e.g., *in situ* hybridization, of the probe/primer to the nucleic acid, the presence or absence of the genetic lesion.

[0259] In preferred embodiments detecting the misexpression includes ascertaining the existence of at least one of: an alteration in the level of a messenger RNA transcript of the 46584 gene; the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; or a non-wild type level of 46584.

[0260] Methods of the invention can be used prenatally or to determine if a subject's offspring will be at risk for a disorder.

[0261] In preferred embodiments the method includes determining the structure of a 46584 gene, an abnormal structure being indicative of risk for the disorder.

[0262] In preferred embodiments the method includes contacting a sample from the subject with an antibody to the

46584 protein or a nucleic acid, which hybridizes specifically with the gene. These and other embodiments are discussed below.

Diagnostic and Prognostic Assays

[0263] The presence, level, or absence of 46584 protein or nucleic acid in a biological sample can be evaluated by obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting 46584 protein or nucleic acid (e.g., mRNA, genomic DNA) that encodes 46584 protein such that the presence of 46584 protein or nucleic acid is detected in the biological sample. The term "biological sample" includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. A preferred biological sample is serum. The level of expression of the 46584 gene can be measured in a number of ways, including, but not limited to: measuring the mRNA encoded by the 46584 genes; measuring the amount of protein encoded by the 46584 genes; or measuring the activity of the protein encoded by the 46584 genes.

[0264] The level of mRNA corresponding to the 46584 gene in a cell can be determined both by *in situ* and by *in vitro* formats.

[0265] The isolated mRNA can be used in hybridization or amplification assays that include, but are not limited to, Southern or Northern analyses, polymerase chain reaction analyses and probe arrays. One preferred diagnostic method for the detection of mRNA levels involves contacting the isolated mRNA with a nucleic acid molecule (probe) that can hybridize to the mRNA encoded by the gene being detected. The nucleic acid probe can be, for example, a full-length 46584 nucleic acid, such as the nucleic acid of SEQ ID NO:1, or a portion thereof, such as an oligonucleotide of at least 7, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to 46584 mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays are described herein.

[0266] In one format, mRNA (or cDNA) is immobilized on a surface and contacted with the probes, for example by running the isolated mRNA on an agarose gel and transferring the mRNA from the gel to a membrane, such as nitrocellulose. In an alternative format, the probes are immobilized on a surface and the mRNA (or cDNA) is contacted with the probes, for example, in a two-dimensional gene chip array. A skilled artisan can adapt known mRNA detection methods for use in detecting the level of mRNA encoded by the 46584 genes.

[0267] The level of mRNA in a sample that is encoded by one of 46584 can be evaluated with nucleic acid amplification, e.g., by rtPCR (Mullis (1987) U.S. Patent No. 4,683,202), ligase chain reaction (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193), self sustained sequence replication (Guatelli *et al.*, (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh *et al.*, (1989), *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi *et al.*, (1988) *Bio/Technology* 6:1197), rolling circle replication (Lizardi *et al.*, U.S. Patent No. 5,854,033) or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques known in the art. As used herein, amplification primers are defined as being a pair of nucleic acid molecules that can anneal to 5' or 3' regions of a gene (plus and minus strands, respectively, or vice-versa) and contain a short region in between. In general, amplification primers are from about 10 to 30 nucleotides in length and flank a region from about 50 to 200 nucleotides in length. Under appropriate conditions and with appropriate reagents, such primers permit the amplification of a nucleic acid molecule comprising the nucleotide sequence flanked by the primers.

[0268] For *in situ* methods, a cell or tissue sample can be prepared/processed and immobilized on a support, typically a glass slide, and then contacted with a probe that can hybridize to mRNA that encodes the 46584 gene being analyzed.

[0269] In another embodiment, the methods further contacting a control sample with a compound or agent capable of detecting 46584 mRNA, or genomic DNA, and comparing the presence of 46584 mRNA or genomic DNA in the control sample with the presence of 46584 mRNA or genomic DNA in the test sample.

[0270] A variety of methods can be used to determine the level of protein encoded by 46584. In general, these methods include contacting an agent that selectively binds to the protein, such as an antibody with a sample, to evaluate the level of protein in the sample. In a preferred embodiment, the antibody bears a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with a detectable substance. Examples of detectable substances are provided herein.

[0271] The detection methods can be used to detect 46584 protein in a biological sample *in vitro* as well as *in vivo*. *In vitro* techniques for detection of 46584 protein include enzyme linked immunosorbent assays (ELISAs), immunoprecipitations, immunofluorescence, enzyme immunoassay (EIA), radioimmunoassay (RIA), and Western blot analysis. *In vivo* techniques for detection of 46584 protein include introducing into a subject a labeled anti-46584 antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

[0272] In another embodiment, the methods further include contacting the control sample with a compound or agent capable of detecting 46584 protein, and comparing the presence of 46584 protein in the control sample with the presence of 46584 protein in the test sample. **[00273]** The invention also includes kits for detecting the presence of 46584 in a biological sample. For example, the kit can include a compound or agent capable of detecting 46584 protein or mRNA in a biological sample; and a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect 46584 protein or nucleic acid.

[0273] For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a marker of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

[0274] For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also include a buffering agent, a preservative, or a protein stabilizing agent. The kit can also include components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

[0275] The diagnostic methods described herein can identify subjects having, or at risk of developing, a disease or disorder associated with misexpressed or aberrant or unwanted 46584 expression or activity. As used herein, the term "unwanted" includes an unwanted phenomenon involved in a biological response such as pain or deregulated cell proliferation.

[0276] In one embodiment, a disease or disorder associated with aberrant or unwanted 46584 expression or activity is identified. A test sample is obtained from a subject and 46584 protein or nucleic acid (e.g., mRNA or genomic DNA) is evaluated, wherein the level, e.g., the presence or absence, of 46584 protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant or unwanted 46584 expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest, including a biological fluid (e.g., serum), cell sample, or tissue.

[0277] The prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant or unwanted 46584 expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a sugar transport disorder.

[0278] The methods of the invention can also be used to detect genetic alterations in a 46584 gene, thereby determining if a subject with the altered gene is at risk for a disorder characterized by misregulation in 46584 protein activity or nucleic acid expression, such as a sugar transport disorder. In preferred embodiments, the methods include detecting, in a sample from the subject, the presence or absence of a genetic alteration characterized by at least one of an alteration affecting the integrity of a gene encoding a 46584-protein, or the mis-expression of the 46584 gene. For example, such genetic alterations can be detected by ascertaining the existence of at least one of 1) a deletion of one or more nucleotides from a 46584 gene; 2) an addition of one or more nucleotides to a 46584 gene; 3) a substitution of one or more nucleotides of a 46584 gene, 4) a chromosomal rearrangement of a 46584 gene; 5) an alteration in the level of a messenger RNA transcript of a 46584 gene, 6) aberrant modification of a 46584 gene, such as of the methylation pattern of the genomic DNA, 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of a 46584 gene, 8) a non-wild type level of a 46584-protein, 9) allelic loss of a 46584 gene, and 10) inappropriate post-translational modification of a 46584-protein.

[0279] An alteration can be detected without a probe/primer in a polymerase chain reaction, such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR), the latter of which can be particularly useful for detecting point mutations in the 46584-gene. This method can include the steps of collecting a sample of cells from a subject, isolating nucleic acid (e.g., genomic, mRNA or both) from the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a 46584 gene under conditions such that hybridization and amplification of the 46584 gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein. Alternatively, other amplification methods described herein or known in the art can be used.

[0280] In another embodiment, mutations in a 46584 gene from a sample cell can be identified by detecting alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined, e.g., by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations

in the sample DNA. Moreover, the use of sequence specific ribozymes (see, for example, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

[0281] In other embodiments, genetic mutations in 46584 can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, two dimensional arrays, e.g., chip based arrays. Such arrays include a plurality of addresses, each of which is positionally distinguishable from the other. A different probe is located at each address of the plurality. The arrays can have a high density of addresses, e.g., can contain hundreds or thousands of oligonucleotide probes (Cronin *et al.* (1996) *Human Mutation* 7: 244-255; Kozal *et al.* (1996) *Nature Medicine* 2: 753-759). For example, genetic mutations in 46584 can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, M.T. *et al. supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

[0282] In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the 46584 gene and detect mutations by comparing the sequence of the sample 46584 with the corresponding wild-type (control) sequence. Automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve *et al.* (1995) *Biotechniques* 19:448-53), including sequencing by mass spectrometry.

[0283] Other methods for detecting mutations in the 46584 gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers *et al.* (1985) *Science* 230:1242; Cotton *et al.* (1988) *Proc. Natl Acad Sci USA* 85:4397; Saleeba *et al.* (1992) *Methods Enzymol.* 217:286-295).

[0284] In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in 46584 cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu *et al.* (1994) *Carcinogenesis* 15:1657-1662; U.S. Patent No. 5,459,039).

[0285] In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in 46584 genes. For example, single strand conformation polymorphism (SSCP) can be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita *et al.* (1989) *Proc Natl. Acad. Sci USA*: 86:2766, see also Cotton (1993) *Mutat. Res.* 285:125-144; and Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control 46584 nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments can be labeled or detected with labeled probes. The sensitivity of the assay can be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen *et al.* (1991) *Trends Genet* 7:5).

[0286] In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers *et al.* (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys Chem* 265:12753).

[0287] Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension (Saiki *et al.* (1986) *Nature* 324:163; Saiki *et al.* (1989) *Proc. Natl Acad. Sci USA* 86:6230).

[0288] Alternatively, allele specific amplification technology which depends on selective PCR amplification can be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification can carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs *et al.* (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini *et al.* (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification can also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci USA* 88:189-93). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

[0289] The methods described herein can be performed, for example, by utilizing prepackaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which can be conveniently used, e.g., in

clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a 46584 gene.

Use of 46584 Molecules as Surrogate Markers

[0290] The 46584 molecules of the invention are also useful as markers of disorders or disease states, as markers for precursors of disease states, as markers for predisposition of disease states, as markers of drug activity, or as markers of the pharmacogenomic profile of a subject. Using the methods described herein, the presence, absence and/or quantity of the 46584 molecules of the invention can be detected, and can be correlated with one or more biological states *in vivo*. For example, the 46584 molecules of the invention can serve as surrogate markers for one or more disorders or disease states or for conditions leading up to disease states. As used herein, a "surrogate marker" is an objective biochemical marker which correlates with the absence or presence of a disease or disorder, or with the progression of a disease or disorder (e.g., with the presence or absence of a tumor). The presence or quantity of such markers is independent of the disease. Therefore, these markers can serve to indicate whether a particular course of treatment is effective in lessening a disease state or disorder. Surrogate markers are of particular use when the presence or extent of a disease state or disorder is difficult to assess through standard methodologies (e.g., early stage tumors), or when an assessment of disease progression is desired before a potentially dangerous clinical endpoint is reached (e.g., an assessment of cardiovascular disease can be made using cholesterol levels as a surrogate marker, and an analysis of HIV infection can be made using HIV RNA levels as a surrogate marker, well in advance of the undesirable clinical outcomes of myocardial infarction or fully-developed AIDS). Examples of the use of surrogate markers in the art include: Koomen *et al.* (2000) *J. Mass. Spectrom.* 35: 258-264; and James (1994) *AIDS Treatment News Archive* 209.

[0291] The 46584 molecules of the invention are also useful as pharmacodynamic markers. As used herein, a "pharmacodynamic marker" is an objective biochemical marker which correlates specifically with drug effects. The presence or quantity of a pharmacodynamic marker is not related to the disease state or disorder for which the drug is being administered; therefore, the presence or quantity of the marker is indicative of the presence or activity of the drug in a subject. For example, a pharmacodynamic marker can be indicative of the concentration of the drug in a biological tissue, in that the marker is either expressed or transcribed or not expressed or transcribed in that tissue in relationship to the level of the drug. In this fashion, the distribution or uptake of the drug can be monitored by the pharmacodynamic marker. Similarly, the presence or quantity of the pharmacodynamic marker can be related to the presence or quantity of the metabolic product of a drug, such that the presence or quantity of the marker is indicative of the relative breakdown rate of the drug *in vivo*. Pharmacodynamic markers are of particular use in increasing the sensitivity of detection of drug effects, particularly when the drug is administered in low doses. Since even a small amount of a drug can be sufficient to activate multiple rounds of marker (e.g., a 46584 marker) transcription or expression, the amplified marker can be in a quantity which is more readily detectable than the drug itself. Also, the marker can be more easily detected due to the nature of the marker itself; for example, using the methods described herein, anti-46584 antibodies can be employed in an immune-based detection system for a 46584 protein marker, or 46584-specific radiolabeled probes can be used to detect a 46584 mRNA marker. Furthermore, the use of a pharmacodynamic marker can offer mechanism-based prediction of risk due to drug treatment beyond the range of possible direct observations. Examples of the use of pharmacodynamic markers in the art include: Matsuda *et al.* US 6,033,862; Hattis *et al.* (1991) *Env. Health Perspect.* 90: 229-238; Schentag (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S21-S24; and Nicolau (1999) *Am. J. Health-Syst. Pharm.* 56 Suppl. 3: S16-S20.

[0292] The 46584 molecules of the invention are also useful as pharmacogenomic markers. As used herein, a "pharmacogenomic marker" is an objective biochemical marker which correlates with a specific clinical drug response or susceptibility in a subject (see, e.g., McLeod *et al.* (1999) *Eur. J. Cancer* 35:1650-1652). The presence or quantity of the pharmacogenomic marker is related to the predicted response of the subject to a specific drug or class of drugs prior to administration of the drug. By assessing the presence or quantity of one or more pharmacogenomic markers in a subject, a drug therapy which is most appropriate for the subject, or which is predicted to have a greater degree of success, can be selected. For example, based on the presence or quantity of RNA, or protein (e.g., 46584 protein or RNA) for specific tumor markers in a subject, a drug or course of treatment can be selected that is optimized for the treatment of the specific tumor likely to be present in the subject. Similarly, the presence or absence of a specific sequence mutation in 46584 DNA can correlate with a 46584 drug response. The use of pharmacogenomic markers therefore permits the application of the most appropriate treatment for each subject without having to administer the therapy.

Pharmaceutical Compositions

[0293] The nucleic acid and polypeptides, fragments thereof, as well as anti-46584 antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions. Such compositions

typically include the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" includes solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Supplementary active compounds can also be incorporated into the compositions.

[0294] A pharmaceutical composition is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

[0295] Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It should be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

[0296] Sterile injectable solutions can be prepared by incorporating the active compound in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

[0297] Oral compositions generally include an inert diluent or an edible carrier. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules, *e.g.*, gelatin capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

[0298] For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

[0299] Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

[0300] The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

[0301] In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be

apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

[0302] It is advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier.

[0303] Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds which exhibit high therapeutic indices are preferred. While compounds that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0304] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[0305] As defined herein, a therapeutically effective amount of protein or polypeptide (i.e., an effective dosage) ranges from about 0.001 to 30 mg/kg body weight, preferably about 0.01 to 25 mg/kg body weight, more preferably about 0.1 to 20 mg/kg body weight, and even more preferably about 1 to 10 mg/kg, 2 to 9 mg/kg, 3 to 8 mg/kg, 4 to 7 mg/kg, or 5 to 6 mg/kg body weight. The protein or polypeptide can be administered one time per week for between about 1 to 10 weeks, preferably between 2 to 8 weeks, more preferably between about 3 to 7 weeks, and even more preferably for about 4, 5, or 6 weeks. The skilled artisan will appreciate that certain factors can influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of a protein, polypeptide, or antibody, unconjugated or conjugated as described herein, can include a single treatment or, preferably, can include a series of treatments.

[0306] For antibodies, the preferred dosage is 0.1 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank *et al.* ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

[0307] The present invention encompasses agents which modulate expression or activity. An agent can, for example, be a small molecule. For example, such small molecules include, but are not limited to, peptides, peptidomimetics (e.g., peptoids), amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

[0308] Exemplary doses include milligram or microgram amounts of the small molecule per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram. It is furthermore understood that appropriate doses of a small molecule depend upon the potency of the small molecule with respect to the expression or activity to be modulated. When one or more of these small molecules is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher can, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific compound employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route

of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

[0309] The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see U. S. Patent 5,328,470) or by stereotactic injection (see e.g., Chen *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g., retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

[0310] The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Methods of Treatment:

[0311] The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant or unwanted 46584 expression or activity. As used herein, the term "treatment" is defined as the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disease, a symptom of disease or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve or affect the disease, the symptoms of disease or the predisposition toward disease. A therapeutic agent includes, but is not limited to, small molecules, peptides, antibodies, ribozymes and antisense oligonucleotides.

[0312] With regards to both prophylactic and therapeutic methods of treatment, such treatments can be specifically tailored or modified, based on knowledge obtained from the field of pharmacogenomics. "Pharmacogenomics", as used herein, refers to the application of genomics technologies such as gene sequencing, statistical genetics, and gene expression analysis to drugs in clinical development and on the market. More specifically, the term refers the study of how a patient's genes determine his or her response to a drug (e.g., a patient's "drug response phenotype", or "drug response genotype".) Thus, another aspect of the invention provides methods for tailoring an individual's prophylactic or therapeutic treatment with either the 46584 molecules of the present invention or 46584 modulators according to that individual's drug response genotype. Pharmacogenomics allows a clinician or physician to target prophylactic or therapeutic treatments to patients who will most benefit from the treatment and to avoid treatment of patients who will experience toxic drug-related side effects.

[0313] In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant or unwanted 46584 expression or activity, by administering to the subject a 46584 or an agent which modulates 46584 expression or at least one 46584 activity. Subjects at risk for a disease which is caused or contributed to by aberrant or unwanted 46584 expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the 46584 aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of 46584 aberrance, for example, a 46584, 46584 agonist or 46584 antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

[0314] It is possible that some 46584 disorders can be caused, at least in part, by an abnormal level of gene product, or by the presence of a gene product exhibiting abnormal activity. As such, the reduction in the level and/or activity of such gene products would bring about the amelioration of disorder symptoms.

[0315] As discussed, successful treatment of 46584 disorders can be brought about by techniques that serve to inhibit the expression or activity of target gene products. For example, compounds, e.g., an agent identified using an assays described above, that proves to exhibit negative modulatory activity, can be used in accordance with the invention to prevent and/or ameliorate symptoms of 46584 disorders. Such molecules can include, but are not limited to peptides, phosphopeptides, small organic or inorganic molecules, or antibodies (including, for example, polyclonal, monoclonal, humanized, human, anti-idiotypic, chimeric or single chain antibodies, and Fab, F(ab')₂ and Fab expression library fragments, scFV molecules, and epitope-binding fragments thereof).

[0316] Further, antisense and ribozyme molecules that inhibit expression of the target gene can also be used in accordance with the invention to reduce the level of target gene expression, thus effectively reducing the level of target gene activity. Still further, triple helix molecules can be utilized in reducing the level of target gene activity. Antisense, ribozyme and triple helix molecules are discussed above.

[0317] It is possible that the use of antisense, ribozyme, and/or triple helix molecules to reduce or inhibit mutant gene expression can also reduce or inhibit the transcription (triple helix) and/or translation (antisense, ribozyme) of mRNA produced by normal target gene alleles, such that the concentration of normal target gene product present can be lower than is necessary for a normal phenotype. In such cases, nucleic acid molecules that encode and express target

gene polypeptides exhibiting normal target gene activity can be introduced into cells via gene therapy method. Alternatively, in instances in that the target gene encodes an extracellular protein, it can be preferable to co-administer normal target gene protein into the cell or tissue in order to maintain the requisite level of cellular or tissue target gene activity.

[0318] Another method by which nucleic acid molecules can be utilized in treating or preventing a disease characterized by 46584 expression is through the use of aptamer molecules specific for 46584 protein. Aptamers are nucleic acid molecules having a tertiary structure which permits them to specifically or selectively bind to protein ligands (see, e.g., Osborne *et al.* (1997) *Curr. Opin. Chem Biol.* 1: 5-9; and Patel (1997) *Curr Opin Chem Biol* 1:32-46). Since nucleic acid molecules can in many cases be more conveniently introduced into target cells than therapeutic protein molecules can be, aptamers offer a method by which 46584 protein activity can be specifically decreased without the introduction of drugs or other molecules which can have pluripotent effects.

[0319] Antibodies can be generated that are both specific for target gene product and that reduce target gene product activity. Such antibodies can, therefore, be administered in instances whereby negative modulatory techniques are appropriate for the treatment of 46584 disorders. For a description of antibodies, see the Antibody section above.

[0320] In circumstances wherein injection of an animal or a human subject with a 46584 protein or epitope for stimulating antibody production is harmful to the subject, it is possible to generate an immune response against 46584 through the use of anti-idiotypic antibodies (see, for example, Herlyn (1999) *Ann Med* 31:66-78; and Bhattacharya-Chatterjee and Foon (1998) *Cancer Treat Res.* 94:51-68). If an anti-idiotypic antibody is introduced into a mammal or human subject, it should stimulate the production of anti-anti-idiotypic antibodies, which should be specific to the 46584 protein. Vaccines directed to a disease characterized by 46584 expression can also be generated in this fashion.

[0321] In instances where the target antigen is intracellular and whole antibodies are used, internalizing antibodies can be preferred. Lipofectin or liposomes can be used to deliver the antibody or a fragment of the Fab region that binds to the target antigen into cells. Where fragments of the antibody are used, the smallest inhibitory fragment that binds to the target antigen is preferred. For example, peptides having an amino acid sequence corresponding to the Fv region of the antibody can be used. Alternatively, single chain neutralizing antibodies that bind to intracellular target antigens can also be administered. Such single chain antibodies can be administered, for example, by expressing nucleotide sequences encoding single-chain antibodies within the target cell population (see e.g., Marasco *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:7889-7893).

[0322] The identified compounds that inhibit target gene expression, synthesis and/or activity can be administered to a patient at therapeutically effective doses to prevent, treat or ameliorate 46584 disorders. A therapeutically effective dose refers to that amount of the compound sufficient to result in amelioration of symptoms of the disorders. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures as described above.

[0323] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[0324] Another example of determination of effective dose for an individual is the ability to directly assay levels of "free" and "bound" compound in the serum of the test subject. Such assays can utilize antibody mimics and/or "biosensors" that have been created through molecular imprinting techniques. The compound which is able to modulate 46584 activity is used as a template, or "imprinting molecule", to spatially organize polymerizable monomers prior to their polymerization with catalytic reagents. The subsequent removal of the imprinted molecule leaves a polymer matrix which contains a repeated "negative image" of the compound and is able to selectively rebinding the molecule under biological assay conditions. A detailed review of this technique can be seen in Ansell *et al* (1996) *Current Opinion in Biotechnology* 7:89-94 and in Shea (1994) *Trends in Polymer Science* 2:166-173. Such "imprinted" affinity matrixes are amenable to ligand-binding assays, whereby the immobilized monoclonal antibody component is replaced by an appropriately imprinted matrix. An example of the use of such matrixes in this way can be seen in Vlatakis *et al* (1993) *Nature* 361:645-647. Through the use of isotope-labeling, the "free" concentration of compound which modulates the expression or activity of 46584 can be readily monitored and used in calculations of IC₅₀.

[0325] Such "imprinted" affinity matrixes can also be designed to include fluorescent groups whose photon-emitting properties measurably change upon local and selective binding of target compound. These changes can be readily assayed in real time using appropriate fiberoptic devices, in turn allowing the dose in a test subject to be quickly optimized based on its individual IC₅₀. An rudimentary example of such a "biosensor" is discussed in Kriz *et al* (1995)

Analytical Chemistry 67:2142-2144.

[0326] Another aspect of the invention pertains to methods of modulating 46584 expression or activity for therapeutic purposes. Accordingly, in an exemplary embodiment, the modulatory method of the invention involves contacting a cell with a 46584 or agent that modulates one or more of the activities of 46584 protein activity associated with the cell. An agent that modulates 46584 protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring target molecule of a 46584 protein (e.g., a 46584 substrate or receptor), a 46584 antibody, a 46584 agonist or antagonist, a peptidomimetic of a 46584 agonist or antagonist, or other small molecule.

[0327] In one embodiment, the agent stimulates one or 46584 activities. Examples of such stimulatory agents include active 46584 protein and a nucleic acid molecule encoding 46584. In another embodiment, the agent inhibits one or more 46584 activities. Examples of such inhibitory agents include antisense 46584 nucleic acid molecules, anti-46584 antibodies, and 46584 inhibitors. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant or unwanted expression or activity of a 46584 protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up regulates or down regulates) 46584 expression or activity. In another embodiment, the method involves administering a 46584 protein or nucleic acid molecule as therapy to compensate for reduced, aberrant, or unwanted 46584 expression or activity.

[0328] Stimulation of 46584 activity is desirable in situations in which 46584 is abnormally downregulated and/or in which increased 46584 activity is likely to have a beneficial effect. For example, stimulation of 46584 activity is desirable in situations in which a 46584 is downregulated and/or in which increased 46584 activity is likely to have a beneficial effect. Likewise, inhibition of 46584 activity is desirable in situations in which 46584 is abnormally upregulated and/or in which decreased 46584 activity is likely to have a beneficial effect.

Pharmacogenomics

[0329] The 46584 molecules of the present invention, as well as agents, or modulators which have a stimulatory or inhibitory effect on 46584 activity (e.g., 46584 gene expression) as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) 46584-associated disorders (e.g., aberrant or deficient sugar transport function or expression) associated with aberrant or unwanted 46584 activity. In conjunction with such treatment, pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) can be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, a physician or clinician can consider applying knowledge obtained in relevant pharmacogenomics studies in determining whether to administer a 46584 molecule or 46584 modulator as well as tailoring the dosage and/or therapeutic regimen of treatment with a 46584 molecule or 46584 modulator.

[0330] Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, for example, Eichelbaum *et al.* (1996) *Clin. Exp. Pharmacol. Physiol.* 23:983-985 and Linder *et al.* (1997) *Clin. Chem.* 43:254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare genetic defects or as naturally-occurring polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (antimalarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

[0331] One pharmacogenomics approach to identifying genes that predict drug response, known as "a genome-wide association", relies primarily on a high-resolution map of the human genome consisting of already known gene-related markers (e.g., a "bi-allelic" gene marker map which consists of 60,000-100,000 polymorphic or variable sites on the human genome, each of which has two variants.) Such a high-resolution genetic map can be compared to a map of the genome of each of a statistically significant number of patients taking part in a Phase II/III drug trial to identify markers associated with a particular observed drug response or side effect. Alternatively, such a high resolution map can be generated from a combination of some ten-million known single nucleotide polymorphisms (SNPs) in the human genome. As used herein, a "SNP" is a common alteration that occurs in a single nucleotide base in a stretch of DNA. For example, a SNP can occur once per every 1000 bases of DNA. A SNP can be involved in a disease process, however, the vast majority can not be disease-associated. Given a genetic map based on the occurrence of such SNPs, individuals can be grouped into genetic categories depending on a particular pattern of SNPs in their individual genome. In such a manner, treatment regimens can be tailored to groups of genetically similar individuals, taking into account traits that can be common among such genetically similar individuals.

[0332] Alternatively, a method termed the "candidate gene approach", can be utilized to identify genes that predict drug response. According to this method, if a gene that encodes a drug's target is known (e.g., a 46584 protein of the present invention), all common variants of that gene can be fairly easily identified in the population and it can be determined if having one version of the gene versus another is associated with a particular drug response.

[0333] Alternatively, a method termed the "gene expression profiling", can be utilized to identify genes that predict drug response. For example, the gene expression of an animal dosed with a drug (e.g., a 46584 molecule or 46584 modulator of the present invention) can give an indication whether gene pathways related to toxicity have been turned on.

[0334] Information generated from more than one of the above pharmacogenomics approaches can be used to determine appropriate dosage and treatment regimens for prophylactic or therapeutic treatment of an individual. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a 46584 molecule or 46584 modulator, such as a modulator identified by one of the exemplary screening assays described herein.

[0335] The present invention further provides methods for identifying new agents, or combinations, that are based on identifying agents that modulate the activity of one or more of the gene products encoded by one or more of the 46584 genes of the present invention, wherein these products can be associated with resistance of the cells to a therapeutic agent. Specifically, the activity of the proteins encoded by the 46584 genes of the present invention can be used as a basis for identifying agents for overcoming agent resistance. By blocking the activity of one or more of the resistance proteins, target cells, e.g., human cells, will become sensitive to treatment with an agent to which the unmodified target cells were resistant.

[0336] Monitoring the influence of agents (e.g., drugs) on the expression or activity of a 46584 protein can be applied in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase 46584 gene expression, protein levels, or upregulate 46584 activity, can be monitored in clinical trials of subjects exhibiting decreased 46584 gene expression, protein levels, or downregulated 46584 activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease 46584 gene expression, protein levels, or downregulate 46584 activity, can be monitored in clinical trials of subjects exhibiting increased 46584 gene expression, protein levels, or upregulated 46584 activity. In such clinical trials, the expression or activity of a 46584 gene, and preferably, other genes that have been implicated in, for example, a COC transporter-associated or another 46584-associated disorder can be used as a "read out" or markers of the phenotype of a particular cell.

Other Embodiments

[0337] In another aspect, the invention features a method of analyzing a plurality of capture probes. The method is useful, e.g., to analyze gene expression. The method includes: providing a two dimensional array having a plurality of addresses, each address of the plurality being positionally distinguishable from each other address of the plurality, and each address of the plurality having a unique capture probe, e.g., a nucleic acid or peptide sequence, wherein the capture probes are from a cell or subject which expresses 46584 or from a cell or subject in which a 46584 mediated response has been elicited; contacting the array with a 46584 nucleic acid (preferably purified), a 46584 polypeptide (preferably purified), or an anti-46584 antibody, and thereby evaluating the plurality of capture probes. Binding, e.g., in the case of a nucleic acid, hybridization with a capture probe at an address of the plurality, is detected, e.g., by a signal generated from a label attached to the 46584 nucleic acid, polypeptide, or antibody.

[0338] The capture probes can be a set of nucleic acids from a selected sample, e.g., a sample of nucleic acids derived from a control or non-stimulated tissue or cell.

[0339] The method can include contacting the 46584 nucleic acid, polypeptide, or antibody with a first array having a plurality of capture probes and a second array having a different plurality of capture probes. The results of each hybridization can be compared, e.g., to analyze differences in expression between a first and second sample. The first plurality of capture probes can be from a control sample, e.g., a wild type, normal, or non-diseased, non-stimulated, sample, e.g., a biological fluid, tissue, or cell sample. The second plurality of capture probes can be from an experimental sample, e.g., a mutant type, at risk, disease-state or disorder-state, or stimulated, sample, e.g., a biological fluid, tissue, or cell sample.

[0340] The plurality of capture probes can be a plurality of nucleic acid probes each of which specifically hybridizes, with an allele of 46584. Such methods can be used to diagnose a subject, e.g., to evaluate risk for a disease or disorder, to evaluate suitability of a selected treatment for a subject, to evaluate whether a subject has a disease or disorder.

[0341] The method can be used to detect SNPs, as described above.

[0342] In another aspect, the invention features, a method of analyzing 46584, e.g., analyzing structure, function, or relatedness to other nucleic acid or amino acid sequences. The method includes: providing a 46584 nucleic acid or amino acid sequence; comparing the 46584 sequence with one or more preferably a plurality of sequences from a collection of sequences, e.g., a nucleic acid or protein sequence database; to thereby analyze 46584.

[0343] The method can include evaluating the sequence identity between a 46584 sequence and a database sequence. The method can be performed by accessing the database at a second site, *e.g.*, over the internet. Preferred databases include GenBank™ and SwissProt.

[0344] In another aspect, the invention features, a set of oligonucleotides, useful, *e.g.*, for identifying SNP's, or identifying specific alleles of 46584. The set includes a plurality of oligonucleotides, each of which has a different nucleotide at an interrogation position, *e.g.*, an SNP or the site of a mutation. In a preferred embodiment, the oligonucleotides of the plurality identical in sequence with one another (except for differences in length). The oligonucleotides can be provided with differential labels, such that an oligonucleotide which hybridizes to one allele provides a signal that is distinguishable from an oligonucleotides which hybridizes to a second aileie.

[0345] The sequences of 46584 molecules are provided in a variety of mediums to facilitate use thereof. A sequence can be provided as a manufacture, other than an isolated nucleic acid or amino acid molecule, which contains a 46584 molecule. Such a manufacture can provide a nucleotide or amino acid sequence, *e.g.*, an open reading frame, in a form which allows examination of the manufacture using means not directly applicable to examining the nucleotide or amino acid sequences, or a subset thereof, as they exist in nature or in purified form.

[0346] A 46584 nucleotide or amino acid sequence can be recorded on computer readable media. As used herein, "computer readable media" refers to any medium that can be read and accessed directly by a computer. Such media include, but are not limited to: magnetic storage media, such as floppy discs, hard disc storage medium, and magnetic tape; optical storage media such as compact disc and CD-ROM; electrical storage media such as RAM, ROM, EPROM, EEPROM, and the like; and general hard disks and hybrids of these categories such as magnetic/optical storage media. The medium is adapted or configured for having thereon 46584 sequence information of the present invention.

[0347] As used herein, the term "electronic apparatus" is intended to include any suitable computing or processing apparatus of other device configured or adapted for storing data or information. Examples of electronic apparatus suitable for use with the present invention include stand-alone computing apparatus; networks, including a local area network (LAN), a wide area network (WAN) Internet, Intranet, and Extranet; electronic appliances such as personal digital assistants (PDAs), cellular phones, pagers, and the like; and local and distributed processing systems.

[0348] As used herein, "recorded" refers to a process for storing or encoding information on the electronic apparatus readable medium. Those skilled in the art can readily adopt any of the presently known methods for recording information on known media to generate manufactures comprising the 46584 sequence information.

[0349] A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a 46584 nucleotide or amino acid sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. The skilled artisan can readily adapt any number of data processor structuring formats (*e.g.*, text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

[0350] By providing the 46584 nucleotide or amino acid sequences of the invention in computer readable form, the skilled artisan can routinely access the sequence information for a variety of purposes. For example, one skilled in the art can use the nucleotide or amino acid sequences of the invention in computer readable form to compare a target sequence or target structural motif with the sequence information stored within the data storage means. A search is used to identify fragments or regions of the sequences of the invention which match a particular target sequence or target motif.

[0351] The present invention therefore provides a medium for holding instructions for performing a method for determining whether a subject has a COC transporter-associated or another 46584-associated disease or disorder or a pre-disposition to a COC transporter-associated or another 46584-associated disease or disorder, wherein the method comprises the steps of determining 46584 sequence information associated with the subject and based on the 46584 sequence information, determining whether the subject has a COC transporter-associated or another 46584-associated disease or disorder and/or recommending a particular treatment for the disease, disorder, or pre-disease condition.

[0352] The present invention further provides in an electronic system and/or in a network, a method for determining whether a subject has a COC transporter-associated or another 46584-associated disease or disorder or a pre-disposition to a disease associated with 46584, wherein the method comprises the steps of determining 46584 sequence information associated with the subject, and based on the 46584 sequence information, determining whether the subject has a COC transporter-associated or another 46584-associated disease or disorder or a pre-disposition to a COC transporter-associated or another 46584-associated disease or disorder, and/or recommending a particular treatment for the disease, disorder, or pre-disease condition. The method may further comprise the step of receiving phenotypic information associated with the subject and/or acquiring from a network phenotypic information associated with the subject.

[0353] The present invention also provides in a network, a method for determining whether a subject has a COC transporter-associated or another 46584-associated disease or disorder or a pre-disposition to a COC transporter-associated or another 46584-associated disease or disorder, said method comprising the steps of receiving 46584 sequence information from the subject and/or information related thereto, receiving phenotypic information associated with the subject, acquiring information from the network corresponding to 46584 and/or corresponding to a COC transporter-associated or another 46584-associated disease or disorder, and based on one or more of the phenotypic information, the 46584 information (e.g., sequence information and/or information related thereto), and the acquired information, determining whether the subject has a COC transporter-associated or another 46584-associated disease or disorder or a pre-disposition to a COC transporter-associated or another 46584-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

[0354] The present invention also provides a business method for determining whether a subject has a COC transporter-associated or another 46584-associated disease or disorder or a pre-disposition to a COC transporter-associated or another 46584-associated disease or disorder, said method comprising the steps of receiving information related to 46584 (e.g., sequence information and/or information related thereto), receiving phenotypic information associated with the subject, acquiring information from the network related to 46584 and/or related to a COC transporter-associated or another 46584-associated disease or disorder, and based on one or more of the phenotypic information, the 46584 information, and the acquired information, determining whether the subject has a COC transporter-associated or another 46584-associated disease or disorder or a pre-disposition to a COC transporter-associated or another 46584-associated disease or disorder. The method may further comprise the step of recommending a particular treatment for the disease, disorder, or pre-disease condition.

[0355] The invention also includes an array comprising a 46584 sequence of the present invention. The array can be used to assay expression of one or more genes in the array. In one embodiment, the array can be used to assay gene expression in a tissue to ascertain tissue specificity of genes in the array. In this manner, up to about 7600 genes can be simultaneously assayed for expression, one of which can be 46584. This allows a profile to be developed showing a battery of genes specifically expressed in one or more tissues.

[0356] In addition to such qualitative information, the invention allows the quantitation of gene expression. Thus, not only tissue specificity, but also the level of expression of a battery of genes in the tissue is ascertainable. Thus, genes can be grouped on the basis of their tissue expression per se and level of expression in that tissue. This is useful, for example, in ascertaining the relationship of gene expression in that tissue. Thus, one tissue can be perturbed and the effect on gene expression in a second tissue can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. In this context, the effect of one cell type on another cell type in response to a biological stimulus can be determined. Such a determination is useful, for example, to know the effect of cell-cell interaction at the level of gene expression. If an agent is administered therapeutically to treat one cell type but has an undesirable effect on another cell type, the invention provides an assay to determine the molecular basis of the undesirable effect and thus provides the opportunity to co-administer a counteracting agent or otherwise treat the undesired effect. Similarly, even within a single cell type, undesirable biological effects can be determined at the molecular level. Thus, the effects of an agent on expression of other than the target gene can be ascertained and counteracted.

[0357] In another embodiment, the array can be used to monitor the time course of expression of one or more genes in the array. This can occur in various biological contexts, as disclosed herein, for example development of a COC transporter-associated or another 46584-associated disease or disorder, progression of COC transporter-associated or another 46584-associated disease or disorder, and processes, such as a cellular transformation associated with the COC transporter-associated or another 46584-associated disease or disorder.

[0358] The array is also useful for ascertaining the effect of the expression of a gene on the expression of other genes in the same cell or in different cells (e.g., ascertaining the effect of 46584 expression on the expression of other genes). This provides, for example, for a selection of alternate molecular targets for therapeutic intervention if the ultimate or downstream target cannot be regulated.

[0359] The array is also useful for ascertaining differential expression patterns of one or more genes in normal and abnormal cells. This provides a battery of genes (e.g., including 46584) that could serve as a molecular target for diagnosis or therapeutic intervention.

[0360] As used herein, a "target sequence" can be any DNA or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. Typical sequence lengths of a target sequence are from about 10 to 100 amino acids or from about 30 to 300 nucleotide residues. However, it is well recognized that commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

[0361] Computer software is publicly available which allows a skilled artisan to access sequence information provided

in a computer readable medium for analysis and comparison to other sequences. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software include, but are not limited to, MacPattern (EMBL), BLASTN and BLASTX (NCBI).

5 **[0362]** Thus, the invention features a method of making a computer readable record of a sequence of a 46584 sequence which includes recording the sequence on a computer readable matrix. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

10 **[0363]** In another aspect, the invention features a method of analyzing a sequence. The method includes: providing a 46584 sequence, or record, in computer readable form; comparing a second sequence to the 46584 sequence; thereby analyzing a sequence. Comparison can include comparing to sequences for sequence identity or determining if one sequence is included within the other, *e.g.*, determining if the 46584 sequence includes a sequence being compared. In a preferred embodiment the 46584 or second sequence is stored on a first computer, *e.g.*, at a first site and the comparison is performed, read, or recorded on a second computer, *e.g.*, at a second site. *E.g.*, the 46584 or second sequence can be stored in a public or proprietary database in one computer, and the results of the comparison performed, read, or recorded on a second computer. In a preferred embodiment the record includes one or more of the following: identification of an ORF; identification of a domain, region, or site; identification of the start of transcription; identification of the transcription terminator; the full length amino acid sequence of the protein, or a mature form thereof; the 5' end of the translated region.

20 **[0364]** The contents of all references, patents and published patent applications cited throughout this application are incorporated herein by reference.

Equivalents

25 **[0365]** Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein.

EP 1 266 906 A1

SEQUENCE LISTING

5 <110> Curtis, Rory A.J.
 <120> 46584, A Human Transporter Family Member
 And Uses Therefor
 10 <130> P71980EP
 <150> 60/298,012
 <151> 2001-06-13
 15 <160> 6
 <170> FastSEQ for Windows Version 4.0
 <210> 1
 <211> 6911
 <212> DNA
 20 <213> Homo sapiens
 <220>
 <221> 5'UTR
 <222> (1)...(194)
 25 <221> CDS
 <222> (5)...(2378)
 <221> 3'UTR
 <222> (2379)...(6911)
 30 <400> 1
 gcgg ccg cgt cga cgc gag ccg gag cgg cgc ccg cac tga ggc ggc tgc 49
 Pro Arg Arg Arg Glu Pro Glu Arg Arg Pro His * Gly Gly Cys
 1 5 10
 35 agt gct gac acc act cag ggc aag ggt gtc cga cgg ctg gag cgt tct 97
 Ser Ala Asp Thr Thr Gln Gly Lys Gly Val Arg Arg Leu Glu Arg Ser
 15 20 25 30
 gtt ttg aac cca aag tgg atg atg ctg tca gag ctg aac cac tga aag 145
 Val Leu Asn Pro Lys Trp Met Met Leu Ser Glu Leu Asn His * Lys
 40 35 40 45
 gag gct gtg aaa att tcc cat ctt ctc att ggc cat cag ttg aga taa 193
 Glu Ala Val Lys Ile Ser His Leu Leu Ile Gly His Gln Leu Arg *
 50 55 60
 45 gat gga aga ctc tta caa gga tag gac ttc act gat gaa ggc tgc caa 241
 Asp Gly Arg Leu Leu Gln Gly * Asp Phe Thr Asp Glu Gly Cys Gln
 65 70 75
 50 gga cat tgc cag aga ggt gaa gaa aca aac agt aaa gaa ggt gaa tca 289
 Gly His Cys Gln Arg Gly Glu Glu Thr Asn Ser Lys Glu Gly Glu Ser
 80 85 90
 agc tgt gga ccg agc cca gga tga ata cac cca gag gtc cta cag tcg 337
 Ser Cys Gly Pro Ser Pro Gly * Ile His Pro Glu Val Leu Gln Ser
 55

EP 1 266 906 A1

	95	100	105	
5	gtt cca aga tga aga aga tga tga tga cta cta ccc ggc tgg aga aac Val Pro Arg * Arg Arg * * * Leu Leu Pro Gly Trp Arg Asn 110 115	385		
10	cta taa tgg tga ggc caa cga tga cga agg ctc aag tga agc cac tga Leu * Trp * Gly Gln Arg * Arg Arg Leu Lys * Ser His * 120 125	433		
15	ggg gca tga tga aga tga tga gat cta tga ggg gga gta tca ggg cat Gly Ala * * Arg * * Asp Leu * Gly Gly Val Ser Gly His 130 135 140	481		
20	ccc cag tat gaa cca agc gaa gga cag cat cgt gtc agt ggg gca gcc Pro Gln Tyr Glu Pro Ser Glu Gly Gln His Arg Val Ser Gly Ala Ala 145 150 155	529		
25	caa ggg cga tga gta caa gga ccg acg gga gct gga atc aga aag gag Gln Gly Arg * Val Gln Gly Pro Thr Gly Ala Gly Ile Arg Lys Glu 160 165 170	577		
30	agc tga cga gga aga gtt ggc cca gca gta tga gct gat aat cca aga Ser * Arg Gly Arg Val Gly Pro Ala Val * Ala Asp Asn Pro Arg 175 180 185	625		
35	atg cgg tca tgg tcg ttt tca gtg ggc cct ttt ctt cgt cct ggg cat Met Arg Ser Trp Ser Phe Ser Val Gly Pro Phe Leu Arg Pro Gly His 190 195 200	673		
40	ggc tct tat ggc aga cgg tgt aga ggt gtt tgt cgt tgg ctt cgt gtt Gly Ser Tyr Gly Arg Arg Cys Arg Gly Val Cys Arg Trp Leu Arg Val 205 210 215	721		
45	acc cag tgc tga gac aga cct ctg cat ccc aaa ttc agg atc tgg atg Thr Gln Cys * Asp Arg Pro Leu His Pro Lys Phe Arg Ile Trp Met 220 225 230	769		
50	gct agg cag cat agt gta cct cgg gat gat ggt ggg ggc gtt ctt ctg Ala Arg Gln His Ser Val Pro Arg Asp Asp Gly Gly Gly Val Leu Leu 235 240 245	817		
55	ggg agg act ggc aga caa agt ggg aag gaa aca gtc tct tct gat ttg Gly Arg Thr Gly Arg Gln Ser Gly Lys Glu Thr Val Ser Ser Asp Leu 250 255 260	865		
60	cat gtc tgt caa cgg att ctt tgc ctt cct ttc ttc att tgt cca agg His Val Cys Gln Arg Ile Leu Cys Leu Pro Phe Phe Ile Cys Pro Arg 265 270 275 280	913		
65	tta tgg ctt ctt tct ctt ctg tcg ctt act ttc tgg att cgg gat tgg Leu Trp Leu Leu Ser Leu Leu Ser Leu Thr Phe Trp Ile Arg Asp Trp 285 290 295	961		
70	agg agc cat acc cac tgt gtt ctc gta ctt tgc tga agt cct ggc ccg Arg Ser His Thr His Cys Val Leu Val Leu Cys * Ser Pro Gly Pro 300 305 310	1009		
75	gga aaa gcg ggg cga aca ctt gag ctg gct ctg cat gtt ctg gat gat Gly Lys Ala Gly Arg Thr Leu Glu Leu Ala Leu His Val Leu Asp Asp 315 320 325	1057		

EP 1 266 906 A1

5	cgg tgg cat cta cgc ctc tgc cat ggc ctg ggc cat cat ccc gca cta	1105
	Arg Trp His Leu Arg Leu Cys His Gly Leu Gly His His Pro Ala Leu	
	330 335 340	
10	cgg gtg gag ctt cag cat ggg atc ggc cta cca gtt tca cag ttg gcg	1153
	Arg Val Glu Leu Gln His Gly Ile Gly Leu Pro Val Ser Gln Leu Ala	
	345 350 355	
15	tgt gtt tgt cat cgt ctg tgc act ccc ctg tgt ctc ctc cgt ggt ggc	1201
	Cys Val Cys His Arg Leu Cys Thr Pro Leu Cys Leu Leu Arg Gly Gly	
	360 365 370 375	
20	cct cac att cat gcc tga aag ccc acg att ctt gtt gga ggt tgg aaa	1249
	Pro His Ile His Ala * Lys Pro Thr Ile Leu Val Gly Gly Trp Lys	
	380 385 390	
25	aca tga tga agc ttg gat gat tct gaa gtt aat tca tga cac caa cat	1297
	Thr * * Ser Leu Asp Asp Ser Glu Val Asn Ser * His Gln His	
	395 400	
30	gag agc ccg ggg tca gcc tga gaa ggt ctt cac ggt aaa caa aat aaa	1345
	Glu Ser Pro Gly Ser Ala * Glu Gly Leu His Gly Lys Gln Asn Lys	
	405 410 415	
35	aac tcc taa aca aat aga tga gct gat tga aat tga gag tga cac agg	1393
	Asn Ser * Thr Asn Arg * Ala Asp * Asn * Glu * His Arg	
	420 425	
40	aac atg gta tag gag gtg ttt tgt tgc gat ccg cac cga gct gta ccg	1441
	Asn Met Val * Glu Val Phe Cys Ser Asp Pro His Arg Ala Val Arg	
	430 435 440	
45	aat ttg gtt gac ttt tat gag atg ttt caa cta ccc agt cag gga taa	1489
	Asn Leu Val Asp Phe Tyr Glu Met Phe Gln Leu Pro Ser Gln Gly *	
	445 450 455	
50	tac aat aaa gct tac aat tgt ttg gtt cac cct gtc ctt tgg gta cta	1537
	Tyr Asn Lys Ala Tyr Asn Cys Leu Val His Pro Val Leu Trp Val Leu	
	460 465 470 475	
55	tgg att atc cgt ttg gtt ccc tga tgt cat taa acc tct gca gtc cga	1585
	Trp Ile Ile Arg Leu Val Pro * Cys His * Thr Ser Ala Val Arg	
	480 485	
60	tga ata tgc att gct aac cag aaa tgt gga gag aga taa ata tgc aaa	1633
	* Ile Cys Ile Ala Asn Gln Lys Cys Gly Glu Arg * Ile Cys Lys	
	490 495 500	
65	ttt cac tat taa ctt tac aat gga aaa tca gat tca tac tgg aat gga	1681
	Phe His Tyr * Leu Tyr Asn Gly Lys Ser Asp Ser Tyr Trp Asn Gly	
	505 510 515	
70	ata cga caa tgg cag att cat agg ggt caa gtt caa atc tgt aac ttt	1729
	Ile Arg Gln Trp Gln Ile His Arg Gly Gln Val Gln Ile Cys Asn Phe	
	520 525 530	
75	caa aga ctc tgt ttt taa gtc ctg cac ctt tga gga tgt aac ttc agt	1777
	Gln Arg Leu Cys Phe * Val Leu His Leu * Gly Cys Asn Phe Ser	
	535 540 545	

EP 1 266 906 A1

	gaa cac cta ctt caa gaa ctg cac att tat tga cac tgt ttt tga caa	1825
	Glu His Leu Leu Gln Glu Leu His Ile Tyr * His Cys Phe * Gln	
	550 555 560	
5	cac aga ttt tga gcc ata taa att cat tga cag tga att taa aaa ctg	1873
	His Arg Phe * Ala Ile * Ile His * Gln * Ile * Lys Leu	
	565 570	
10	ctc gtt ttt tca caa caa gac ggg atg tca gat tac ctt tga tga tga	1921
	Leu Val Phe Ser Gln Gln Asp Gly Met Ser Asp Tyr Leu * * *	
	575 580 585	
15	cta tag tgc cta ctg gat tta ttt tgt caa ctt tct ggg gac att ggc	1969
	Leu * Cys Leu Leu Asp Leu Phe Cys Gln Leu Ser Gly Asp Ile Gly	
	590 595 600	
20	agt att gcc agg gaa cat tgt gtc tgc tct gct gat gga cag aat tgg	2017
	Ser Ile Ala Arg Glu His Cys Val Cys Ser Ala Asp Gly Gln Asn Trp	
	605 610 615	
25	gcg ctt aac aat gct agg tgg ctc tat ggt gct ttc ggg gat cag ctg	2065
	Ala Leu Asn Asn Ala Arg Trp Leu Tyr Gly Ala Phe Gly Asp Gln Leu	
	620 625 630	
30	ttt ctt cct ttg gtt cgg cac cag tga atc cat gat gat agg cat gct	2113
	Phe Leu Pro Leu Val Arg His Gln * Ile His Asp Asp Arg His Ala	
	635 640 645	
35	gtg tct gta caa tgg att gac cat ctc agc ctg gaa ctc tct tga cgt	2161
	Val Ser Val Gln Trp Ile Asp His Leu Ser Leu Glu Leu Ser * Arg	
	650 655 660	
40	ggt cac tgt gga act gta ccc cac aga ccg gag ggc aac agg ctt tgg	2209
	Gly His Cys Gly Thr Val Pro His Arg Pro Glu Gly Asn Arg Leu Trp	
	665 670 675	
45	ctt ctt aaa tgc gct atg caa ggc agc agc cgt cct ggg aaa ctt aat	2257
	Leu Leu Lys Cys Ala Met Gln Gly Ser Ser Arg Pro Gly Lys Leu Asn	
	680 685 690 695	
50	att tgg ctc tct ggt cag cat cac caa atc aat ccc cat cct gct ggc	2305
	Ile Trp Leu Ser Gly Gln His His Gln Ile Asn Pro His Pro Ala Gly	
	700 705 710	
55	ttc tac tgt gct cgt gtg tgg agg act cgt tgg gct gtg cct gcc tga	2353
	Phe Tyr Cys Ala Arg Val Trp Arg Thr Arg Trp Ala Val Pro Ala *	
	715 720 725	
60	cac acg aac cca ggt tct gat gta a tgggaaaaaa agccatcctt	2398
	His Thr Asn Pro Gly Ser Asp Val	
	730	
65	cctgcgtttc ttctctctgc cctgggtcaa ttctccttcc tgactcaagg cttcagagtt	2458
	ttcclatata gaaaggtgat caagtatcag aacataaaca cgtgctgtga cttaaaaattt	2518
	agaagcatat catcttgccc ctttgtgatt ttgcacaggt tgtttgtttg ttttgttttg	2578
	tttgaatgca ttgttatctt tccaaactgt gatgctactg cctcccgcaa atcagtgga	2638
	gcattttctaa aatgccctag gcagccaggc ttccctgggg ttctgctcat tgaaaagtta	2698
	gaagctaaga agactgggct ggtgtaagaa atagatcctg ctcttaattt aacattaaca	2758
	ggaaatataa gtcggcacat tattgcattt gtgctgagaa gagggattct tttttttttt	2818
	ttttctaaca gagtgatatt ccctgtttac ttagaaaagg acaccagta attctcactg	2878
	ttatagcaga gcctttcaaa gaaaaccatg tggcaccaaa cagggctggg tggggctctt	2938

	tggggcaaga	caattaaggt	gaccgctgac	aagaaaaata	aaagctctgc	aattagcagc	2998
	aaaaatctga	gtgtgctgca	taagcaagat	ttctgtcaga	aagcccaatt	aaacctctga	3058
	aaagtgtgag	gaaaaagaac	tgtaacagta	tctgtgactg	tggcatgtgg	ttcagaatgt	3118
5	ttgaaaatca	gagatccgag	agagagcttc	tgccctccatt	taccactccc	actccttcag	3178
	ctgcccttgg	tcttgccaga	cagcagattc	agggaaagaa	agctgcttct	taaaaaactgt	3238
	ctgcctacct	gattctgcca	cttaccocaca	ctgtctctag	tctagagatt	gccagtggtg	3298
	agctaaattg	taaattgcaa	attaggaaga	aattcacacc	cagagccatt	tggaatatca	3358
	aaatgggtacc	tctttccagt	acctctttgc	tgtgggactg	gcccgcagg	acccccacc	3418
	cacctggccc	gcagttgaca	ttcagcttcc	cagctggcat	cagcaaaagag	ggtttgtccc	3478
10	tctagcttcc	ccagggctgg	ccactgctc	aatgcgctcc	tggctccctg	taaaaatcaca	3538
	gcccgcctgc	actcctgagg	gcccctctca	cagacagacc	atctccctgt	ggttgtttga	3598
	ctgctcattt	tacctctgga	atctgcctgg	gcttggaaga	aagaggtcag	ccaggacatt	3658
	cccacggggc	cgctgtcage	tacatgacct	tctgcgctct	caggctcccc	ttccccagg	3718
	ctcgtccttt	ttacacctct	tctttgaaat	ctgaaccttc	cttgtagact	tgacgccaat	3778
	gaccagaagc	caggaataaa	tcccatattt	aaacaatatt	cattttttta	ctgcctcgat	3838
15	agaggtttat	tcttgttact	tattataggg	attttcaagc	aagccatagc	ttggggctcc	3898
	caggaaggga	aagcttttgt	ttccactgat	cattattttc	agagatgtcc	taactccctc	3958
	gaggccccc	tgccctcccc	tccccttcc	aatgatccct	cgcagagctc	agactaagtg	4018
	aaaagcttct	ttcattaaca	taagccatta	atcgacttcc	ccttgaagta	attcagggtc	4078
	gcagaggggg	gaagactggg	gttgggggca	gtgggaagag	gggtggtccc	tgggggtctg	4138
20	tgatcagggc	tagaacatca	ggaaaacccg	aacaaaaaac	tgcaaaagtg	gcaggaagaa	4198
	gaaaaatcac	tgaaggttag	acaagatgca	tttgaaagat	accaaccgca	cagggaagaa	4258
	aacaaacttt	tctcacttct	gcctcctatc	attttatttg	agaacaaaga	cagcagtcaa	4318
	acaacccaag	aaagtacctt	gtatgttttt	cttcttccac	atgtaccat	tttccacact	4378
	atctctccct	gatttgtctt	ctttaaaaag	acatttgagg	acctgagcag	gtcttttagt	4438
	tgaatgcagc	atgtgcagaa	ataactgcaa	aagccaggag	tgcatctcca	tcaacccatc	4498
25	aaaagaaaaa	tagaaatgta	catgggcata	gctgtctcta	gtcctctgtg	atagttcact	4558
	ttgctaacta	gactaccttt	acttttcaaa	atatctctaa	tgtctgctct	ggggtctaaa	4618
	tgcataaaa	atctcccaag	ctgcagagca	cagttattca	gaattttatt	tgaacactgg	4678
	gttggttcag	cataggataa	accagcaaaa	ccttacttca	tgagacctgc	catcgtgttc	4738
	ccctgcagaa	ggaccgttct	tgccctactt	gttaccta	ccaagcatcc	ttggtggcct	4798
	aagaggggat	ttttctcatc	tcctggacag	tctgtggtca	caaaccttct	tggtcaccac	4858
30	ctctcccacc	ttttgtgcac	tcctccctgc	ctccttgaga	ctgttcatca	ccttgatttt	4918
	ctaagagagc	tcaaatcatt	tcctaaagcat	gtggtgtggc	ctgtatagct	tgctccagat	4978
	ttagctggag	attaagaaa	agccaggata	ttaaacagaa	actctcagtt	gtatcagttg	5038
	agcaggccac	tggcaccta	acacgcaact	ggcatctcat	ttcgccata	gtagaaggaa	5098
	aatgaaaatt	tgtacaaaaa	tgaaaaccta	ggaaatagtc	cacttgctcc	tatattttta	5158
	ccaccttttc	taactaaaa	ttccctgttg	ctctttggaa	tcaatccaga	atgctcccaa	5218
35	gtccattat	tatcacagaa	tcttctcagt	agggaaaaac	aaacaactgc	ttcaaaactgt	5278
	gaattctgta	tgtaccacat	tctatcccag	gtattgggtt	ttaaggcacc	caactcacat	5338
	tcagtctctg	tgccctgaatc	atctcacacc	agagtacaga	ccttgagagc	gcgatagatg	5398
	ggtcctagcc	tgaccttcac	cactaagcgg	tacagggacc	cagatcttgc	tgtggagact	5458
	gtaagtctgt	gaacatctct	agttctcact	ttccccgtgc	acaaaaataa	ggattgggac	5518
	tagtctgctc	tccactctcc	agcattgtag	aagacagtaa	tgaaaatagc	ttcatcttat	5578
40	ctgccttctt	atgctaaaact	catttttttt	tttttttaggt	ggagtttcac	tctgttgccc	5638
	aggctggagt	gcagtggtgg	gaactctgct	cactgcaacc	tccgcctccc	aggttcaagt	5698
	gattctccca	cctcagcctc	ccaagtagct	gggaccacac	gcattgcgca	ccacgcccag	5758
	ctaatttttg	tatcttttag	tagagggtgg	gtttcaccgt	gttggtcagg	ctggtctcga	5818
	actcctgacc	tcaagtgate	ggcccacctc	agcctcccaa	aatgctggga	ttacaggcgt	5878
45	gagccactgg	cgccctcaac	tcattttttac	aagcactaaa	taggtttagt	ttataaaactg	5938
	aaaaattagg	aaataggcac	gcataaaaag	agtgtaggaa	gttggttggg	aagcccgcat	5998
	ggaataggcc	agaaagcaca	gtccccatgc	tgtctgccct	gccagcatct	cagcatcctg	6058
	actagtatct	ctgttccactg	caggtttgac	agttaccatc	accagccag	aaaaggggccc	6118
	cgccttgcca	gataaacgcg	cttaggttag	aagactataa	ttgcaatatt	acctacagc	6178
	tctcctgact	gtttattcca	ttattttctt	gaaacatggc	cttaagtatt	atacataggc	6238
50	ttcagaaaca	gtttttcagt	gattttacct	atattggcaa	atagacaaat	ctcaccaaat	6298
	cagagttcag	gttttcacca	aaaactgctt	agttgtcttt	caagatacct	taagagatca	6358
	aaggggaggt	gggaaaagaa	atgtagtgc	agaagggaga	gtccaagata	ccactgataa	6418
	ggttccagca	aagagattag	gacacagcag	agactttcta	aatgctgggg	tcattttccaa	6478
	agcagtcag	gggaataact	tctgattccc	tctgccaatc	aaaatagata	gaatatgata	6538
55	ttctccaatt	ttttttctgt	ttgatgtcaa	aatggccttt	acctatttct	tcagtaatat	6598

EP 1 266 906 A1

5 acctgatctg ttcatgcact aaaggtctctg gagccaggaa tgtctgaatt tggccttcac 6658
 ctgggtatcac agaaagacaa ctgggatatt ggtgttaagg ggcataatcgg ctgtatgtta 6718
 gggttttcttt aatcctcaca tcgtgccaaa cttagtctgg ttactaagcc taaaaacacc 6778
 agtattgcaa caacttctga aactctgaag tggactatat gtttctcaca gtttttgttt 6838
 atgctcagca aagtatgtat ttttaagtaaa gcttctcttg ttatctgtaa tgtagacaaa 6898
 aaaaaaaaaa agg 6911

10 <210> 2
 <211> 734
 <212> PRT
 <213> Homo sapiens

<400> 2
 15 Pro Arg Arg Arg Glu Pro Glu Arg Arg Pro His Gly Gly Cys Ser Ala
 1 5 10 15
 Asp Thr Thr Gln Gly Lys Gly Val Arg Arg Leu Glu Arg Ser Val Leu
 20 20 25 30
 Asn Pro Lys Trp Met Met Leu Ser Glu Leu Asn His Lys Glu Ala Val
 35 40 45
 Lys Ile Ser His Leu Leu Ile Gly His Gln Leu Arg Asp Gly Arg Leu
 50 55 60
 20 Leu Gln Gly Asp Phe Thr Asp Glu Gly Cys Gln Gly His Cys Gln Arg
 65 70 75 80
 Gly Glu Glu Thr Asn Ser Lys Glu Gly Glu Ser Ser Cys Gly Pro Ser
 85 90 95
 Pro Gly Ile His Pro Glu Val Leu Gln Ser Val Pro Arg Arg Arg Leu
 100 105 110
 25 Leu Pro Gly Trp Arg Asn Leu Trp Gly Gln Arg Arg Arg Leu Lys Ser
 115 120 125
 His Gly Ala Arg Asp Leu Gly Gly Val Ser Gly His Pro Gln Tyr Glu
 130 135 140
 Pro Ser Glu Gly Gln His Arg Val Ser Gly Ala Ala Gln Gly Arg Val
 145 150 155 160
 30 Gln Gly Pro Thr Gly Ala Gly Ile Arg Lys Glu Ser Arg Gly Arg Val
 165 170 175
 Gly Pro Ala Val Ala Asp Asn Pro Arg Met Arg Ser Trp Ser Phe Ser
 180 185 190
 Val Gly Pro Phe Leu Arg Pro Gly His Gly Ser Tyr Gly Arg Arg Cys
 195 200 205
 35 Arg Gly Val Cys Arg Trp Leu Arg Val Thr Gln Cys Asp Arg Pro Leu
 210 215 220
 His Pro Lys Phe Arg Ile Trp Met Ala Arg Gln His Ser Val Pro Arg
 225 230 235 240
 Asp Asp Gly Gly Gly Val Leu Leu Gly Arg Thr Gly Arg Gln Ser Gly
 245 250 255
 40 Lys Glu Thr Val Ser Ser Asp Leu His Val Cys Gln Arg Ile Leu Cys
 260 265 270
 Leu Pro Phe Ile Cys Pro Arg Leu Trp Leu Leu Ser Leu Leu Ser
 275 280 285
 45 Leu Thr Phe Trp Ile Arg Asp Trp Arg Ser His Thr His Cys Val Leu
 290 295 300
 Val Leu Cys Ser Pro Gly Pro Gly Lys Ala Gly Arg Thr Leu Glu Leu
 305 310 315 320
 Ala Leu His Val Leu Asp Asp Arg Trp His Leu Arg Leu Cys His Gly
 325 330 335
 50 Leu Gly His His Pro Ala Leu Arg Val Glu Leu Gln His Gly Ile Gly
 340 345 350
 Leu Pro Val Ser Gln Leu Ala Cys Val Cys His Arg Leu Cys Thr Pro
 355 360 365
 Leu Cys Leu Leu Arg Gly Gly Pro His Ile His Ala Lys Pro Thr Ile
 370 375 380

55

EP 1 266 906 A1

Leu Val Gly Gly Trp Lys Thr Ser Leu Asp Asp Ser Glu Val Asn Ser
 385 390 395 400
 His Gln His Glu Ser Pro Gly Ser Ala Glu Gly Leu His Gly Lys Gln
 405 410 415
 Asn Lys Asn Ser Thr Asn Arg Ala Asp Asn Glu His Arg Asn Met Val
 420 425 430
 Glu Val Phe Cys Ser Asp Pro His Arg Ala Val Arg Asn Leu Val Asp
 435 440 445
 Phe Tyr Glu Met Phe Gln Leu Pro Ser Gln Gly Tyr Asn Lys Ala Tyr
 450 455 460
 Asn Cys Leu Val His Pro Val Leu Trp Val Leu Trp Ile Ile Arg Leu
 465 470 475 480
 Val Pro Cys His Thr Ser Ala Val Arg Ile Cys Ile Ala Asn Gln Lys
 485 490 495
 Cys Gly Glu Arg Ile Cys Lys Phe His Tyr Leu Tyr Asn Gly Lys Ser
 500 505 510
 Asp Ser Tyr Trp Asn Gly Ile Arg Gln Trp Gln Ile His Arg Gly Gln
 515 520 525
 Val Gln Ile Cys Asn Phe Gln Arg Leu Cys Phe Val Leu His Leu Gly
 530 535 540
 Cys Asn Phe Ser Glu His Leu Leu Gln Glu Leu His Ile Tyr His Cys
 545 550 555 560
 Phe Gln His Arg Phe Ala Ile Ile His Gln Ile Lys Leu Leu Val Phe
 565 570 575
 Ser Gln Gln Asp Gly Met Ser Asp Tyr Leu Leu Cys Leu Leu Asp Leu
 580 585 590
 Phe Cys Gln Leu Ser Gly Asp Ile Gly Ser Ile Ala Arg Glu His Cys
 595 600 605
 Val Cys Ser Ala Asp Gly Gln Asn Trp Ala Leu Asn Asn Ala Arg Trp
 610 615 620
 Leu Tyr Gly Ala Phe Gly Asp Gln Leu Phe Leu Pro Leu Val Arg His
 625 630 635 640
 Gln Ile His Asp Asp Arg His Ala Val Ser Val Gln Trp Ile Asp His
 645 650 655
 Leu Ser Leu Glu Leu Ser Arg Gly His Cys Gly Thr Val Pro His Arg
 660 665 670
 Pro Glu Gly Asn Arg Leu Trp Leu Leu Lys Cys Ala Met Gln Gly Ser
 675 680 685
 Ser Arg Pro Gly Lys Leu Asn Ile Trp Leu Ser Gly Gln His His Gln
 690 695 700
 Ile Asn Pro His Pro Ala Gly Phe Tyr Cys Ala Arg Val Trp Arg Thr
 705 710 715 720
 Arg Trp Ala Val Pro Ala His Thr Asn Pro Gly Ser Asp Val
 725 730

 <210> 3
 <211> 2181
 <212> DNA
 <213> Homo sapiens

 <400> 3
 atggaagact cttacaagga taggacttca ctgatgaagg gtgccaaagga cattgccaga 60
 gaggtgaaga aacaaacagt aaagaagggtg aatcaagctg tggaccgagc ccaggatgaa 120
 tacaccaga ggtcctacag tcggttccaa gatgaagaag atgatgatga ctactaccg 180
 gctggagaaa cctataatgg tgaggccaac gatgacgaag gctcaagtga agccactgag 240
 gggcatgatg aagatgatga gatctatgag ggggagtatc agggcatccc cagtatgaac 300
 caagcgaagg acagcatcgt gtcagtgggg cagcccaagg gcgatgagta caaggaccga 360
 cgggagctgg aatcagaaag gagagctgac gaggaagagt tggcccagca gtatgagctg 420
 ataatccaag aatgcggtca tggtcgtttt cagtgggccc ttttcttcgt cctgggcatg 480
 gctcttatgg cagacggtgt agaggtgttt gtcgttggtc tcgtgttacc cagtgtctgag 540

EP 1 266 906 A1

```

acagacctct gcatcccaaa ttcaggatct ggatggctag gcagcatagt gtacctcggg 600
atgatgggtg gggcggtctt ctggggagga ctggcagaca aagtgggaag gaaacagtct 660
cttctgattt gcatgtctgt caacggattc tttgccttcc tttcttcatt tgtccaaggt 720
tatggcttct ttctcttctg tcgcttactt tctggattcg ggattggagg agccataccc 780
actgtgttct cgtactttgc tgaagtcctg gcccgggaaa agcggggcga acacttgagc 840
tggctctgca tgttctggat gatcgggtgc atctacgcct ctgccatggc ctgggccatc 900
atcccgcact acgggtggag cttcagcatg ggatcggcct accagtttca cagttggcgt 960
gtgtttgtca tcgtctgtgc actccctctg gtctcctccg tgggtggcct caccattcatg 1020
cctgaaagcc cactgattctt gttggagggt ggaaaacatg atgaagcttg gatgattctg 1080
aagttaattc atgacaccaa catgagagcc cggggtcagc ctgagaaggt cttcacggta 1140
aacaaaataa aaactcctaa acaaatagat gagctgattg aaattgagag tgacacagga 1200
acatggtata ggaggtgttt tgttcggatc cgcaccgagc tgtacggaat ttggttgact 1260
tttatgagat gtttcaacta cccagtcagg gataatacaa taaagcttac aattgttttg 1320
ttcacctgt cttttgggta ctatggatta tccgttttgt tccctgatgt cattaacact 1380
ctgcagtcct atgaatatgc attgctaacc agaaatgttg agagagataa atatgcaaat 1440
ttcactatta actttacaat ggaaaatcag attcatactg gaatggaata cgacaatggc 1500
agattcatag ggggtcaagtt caaatctgta actttcaaag actctgtttt taagtcttgc 1560
acctttgagg atgtaacttc agtgaacacc tacttcaaga actgcacatt tattgacact 1620
gtttttgaca acacagattt tgagccatat aaattcattg acagtgaatt taaaactgc 1680
tcgttttttc acaacaagac gggatgtcag attacctttg atgatgacta tagtgccctac 1740
tggatttatt ttgtcaactt tctggggaca ttggcagtat tgccaggga cattgtgtct 1800
gctctgtcga tggacagaat tggcgctta acaatgctag gtggctctat ggtgctttcg 1860
gggaltcagct gtttcttctt ttggttcggc accagtgaat ccatgatgat aggcactgtg 1920
tgtctgtaca atggattgac catctcagcc tggaaactctc ttgacgttgt cactgtggaa 1980
ctgtaccca cagaccggag ggcaacaggc tttggcttct taaatgcgct atgcaaggca 2040
gcagecgtcc tgggaaactt aatatttggc tctctgtgca gcatcaccaa atcaatcccc 2100
atcctgctgg cttctactgt gctcgtgtgt ggaggactcg ttgggctgtg cctgcctgac 2160
acacgaaccc aggttctgat g
2181

```

<210> 4

<211> 727

<212> PRT

<213> Rattus norvegicus

<400> 4

```

Met Glu Asp Ser Tyr Lys Asp Arg Thr Ser Leu Met Lys Gly Ala Lys
 1           5           10          15
Asp Ile Ala Lys Glu Val Lys Lys Gln Thr Val Lys Lys Val Asn Gln
20          25          30
Ala Val Asp Arg Ala Gln Asp Glu Tyr Thr Gln Arg Ser Tyr Ser Arg
35          40          45
Phe Gln Asp Glu Asp Asp Asp Asp Tyr Tyr Pro Pro Gly Glu Thr
50          55          60
Tyr Ser Gly Glu Ala Asn Asp Asp Glu Gly Ser Ser Glu Ala Thr Glu
40 65          70          75          80
Gly His Asp Glu Glu Asp Glu Ile Tyr Glu Gly Glu Tyr Gln Gly Ile
85          90          95
Pro Ser Thr Asn Gln Gly Lys Asp Ser Ile Val Ser Val Gly Gln Pro
100         105         110
Lys Gly Asp Glu Tyr Lys Asp Arg Arg Glu Leu Glu Ser Glu Arg Arg
45 115         120         125
Ala Asp Glu Glu Glu Leu Ala Gln Gln Tyr Glu Leu Ile Ile Gln Glu
130         135         140
Cys Gly His Gly Arg Phe Gln Trp Ala Leu Phe Phe Val Leu Gly Met
145         150         155         160
Ala Leu Met Ala Asp Gly Val Glu Val Phe Val Val Gly Phe Val Leu
50 165         170         175
Pro Ser Ala Glu Thr Asp Leu Cys Ile Pro Asn Ser Gly Ser Gly Trp
180         185         190
Leu Gly Ser Ile Val Tyr Leu Gly Met Met Val Gly Ala Phe Phe Trp
195         200         205

```

EP 1 266 906 A1

	Gly	Gly	Leu	Ala	Asp	Lys	Val	Gly	Arg	Lys	Gln	Ser	Leu	Leu	Ile	Cys
	210						215					220				
5	Met	Ser	Val	Asn	Gly	Phe	Phe	Ala	Phe	Leu	Ser	Ser	Phe	Val	Gln	Gly
	225				230						235					240
	Tyr	Gly	Phe	Phe	Leu	Leu	Cys	Arg	Leu	Leu	Ser	Gly	Phe	Gly	Ile	Gly
					245						250				255	
	Gly	Ala	Ile	Pro	Thr	Val	Phe	Ser	Tyr	Phe	Ala	Glu	Val	Leu	Ala	Arg
				260					265					270		
10	Glu	Lys	Arg	Gly	Glu	His	Leu	Ser	Trp	Leu	Cys	Met	Phe	Trp	Met	Ile
		275						280					285			
	Gly	Gly	Ile	Tyr	Ala	Ser	Ala	Met	Ala	Trp	Ala	Ile	Ile	Pro	His	Tyr
		290					295					300				
	Gly	Trp	Ser	Phe	Ser	Met	Gly	Ser	Ala	Tyr	Gln	Phe	His	Ser	Trp	Arg
	305					310					315					320
15	Val	Phe	Val	Ile	Val	Cys	Ala	Leu	Pro	Cys	Val	Ser	Ser	Val	Val	Ala
					325					330					335	
	Leu	Thr	Phe	Met	Pro	Glu	Ser	Pro	Arg	Phe	Leu	Leu	Glu	Val	Gly	Lys
				340					345					350		
	His	Asp	Glu	Ala	Trp	Met	Ile	Leu	Lys	Leu	Ile	His	Asp	Thr	Asn	Met
		355						360					365			
20	Arg	Ala	Arg	Gly	Gln	Pro	Glu	Lys	Val	Phe	Thr	Val	Asn	Lys	Ile	Lys
		370					375					380				
	Thr	Pro	Lys	Gln	Ile	Asp	Glu	Leu	Ile	Glu	Ile	Glu	Ser	Asp	Thr	Gly
	385					390					395					400
	Thr	Trp	Tyr	Arg	Arg	Cys	Phe	Val	Arg	Ile	Arg	Thr	Glu	Leu	Tyr	Gly
				405					410						415	
25	Ile	Trp	Leu	Thr	Phe	Met	Arg	Cys	Phe	Asn	Tyr	Pro	Val	Arg	Glu	Asn
			420						425					430		
	Thr	Ile	Lys	Leu	Thr	Ile	Val	Trp	Phe	Thr	Leu	Ser	Phe	Gly	Tyr	Tyr
		435					440						445			
	Gly	Leu	Ser	Val	Trp	Phe	Pro	Asp	Val	Ile	Lys	His	Leu	Gln	Ser	Asp
		450					455				460					
30	Glu	Tyr	Ala	Leu	Leu	Thr	Arg	Asn	Val	Gln	Lys	Asp	Lys	Tyr	Ala	Asn
	465					470					475					480
	Phe	Ser	Ile	Asn	Phe	Thr	Met	Glu	Asn	Gln	Val	His	Thr	Gly	Met	Glu
				485					490					495		
	Tyr	Asp	Asn	Gly	Arg	Phe	Leu	Gly	Val	Lys	Phe	Lys	Ser	Val	Thr	Phe
			500						505					510		
35	Lys	Asp	Ser	Val	Phe	Lys	Ser	Cys	Thr	Phe	Asp	Asp	Val	Thr	Ser	Val
		515						520					525			
	Asn	Thr	Tyr	Phe	Lys	Asn	Cys	Thr	Phe	Ile	Asp	Thr	Leu	Phe	Glu	Asn
		530					535					540				
	Thr	Asp	Phe	Glu	Pro	Tyr	Lys	Phe	Ile	Asp	Ser	Glu	Phe	Gln	Asn	Cys
	545					550					555					560
40	Ser	Phe	Leu	His	Asn	Lys	Thr	Gly	Cys	Gln	Ile	Thr	Phe	Asp	Asp	Asp
				565					570						575	
	Tyr	Ser	Ala	Tyr	Trp	Ile	Tyr	Phe	Val	Asn	Phe	Leu	Gly	Thr	Leu	Ala
			580						585					590		
	Val	Leu	Pro	Gly	Asn	Ile	Val	Ser	Ala	Leu	Leu	Met	Asp	Arg	Ile	Gly
		595					600						605			
45	Arg	Leu	Thr	Met	Leu	Gly	Gly	Ser	Met	Val	Leu	Ser	Gly	Ile	Ser	Cys
		610				615						620				
	Phe	Phe	Leu	Trp	Phe	Gly	Thr	Ser	Glu	Ser	Met	Met	Ile	Gly	Met	Leu
	625					630					635					640
	Cys	Leu	Tyr	Asn	Gly	Leu	Thr	Ile	Ser	Ala	Trp	Asn	Ser	Leu	Asp	Val
				645					650						655	
50	Val	Thr	Val	Glu	Leu	Tyr	Pro	Thr	Asp	Arg	Arg	Ala	Thr	Gly	Phe	Gly
			660						665					670		
	Phe	Leu	Asn	Ala	Leu	Cys	Lys	Ala	Ala	Ala	Val	Leu	Gly	Asn	Leu	Ile
		675					680					685				
55	Phe	Gly	Ser	Leu	Val	Ser	Ile	Thr	Lys	Ala	Ile	Pro	Ile	Leu	Leu	Ala

EP 1 266 906 A1

```

        690              695              700
Ser Thr Val Leu Val Cys Gly Gly Leu Val Gly Leu Arg Leu Pro Asp
705              710              715              720
5  Thr Arg Thr Gln Val Leu Met
        725

<210> 5
<211> 488
10 <212> PRT
    <213> Artificial Sequence

<220>
<223> Consensus sequence for PFAM PF00083

15 <400> 5
    Val Ala Leu Val Ala Ala Leu Gly Gly Gly Phe Leu Phe Gly Tyr Asp
        1              5              10              15
    Thr Gly Val Ile Gly Gly Phe Leu Ala Leu Ile Asp Phe Leu Phe Arg
        20              25              30
    Phe Gly Leu Leu Thr Ser Ser Gly Ala Leu Ala Glu Leu Val Gly Tyr
        35              40              45
20 Ser Thr Val Leu Thr Gly Leu Val Val Ser Ile Phe Phe Leu Gly Arg
        50              55              60
    Leu Ile Gly Ser Leu Phe Ala Gly Lys Leu Gly Asp Arg Phe Gly Arg
        65              70              75              80
    Lys Lys Ser Leu Leu Ile Ala Leu Val Leu Phe Val Ile Gly Ala Leu
        85              90              95
25 Leu Ser Gly Ala Ala Pro Gly Tyr Thr Thr Ile Gly Leu Trp Ala Phe
        100              105              110
    Tyr Leu Leu Ile Val Gly Arg Val Leu Val Gly Leu Gly Val Gly Gly
        115              120              125
30 Ala Ser Val Leu Val Pro Met Tyr Ile Ser Glu Ile Ala Pro Lys Ala
        130              135              140
    Leu Arg Gly Ala Leu Gly Ser Leu Tyr Gln Leu Ala Ile Thr Ile Gly
        145              150              155              160
    Ile Leu Val Ala Ala Ile Ile Gly Leu Gly Leu Asn Lys Thr Asn Asn
        165              170              175
35 Asp Ser Ala Leu Asn Ser Trp Gly Trp Arg Ile Pro Leu Gly Leu Gln
        180              185              190
    Leu Val Pro Ala Leu Leu Leu Leu Ile Gly Leu Leu Phe Leu Pro Glu
        195              200              205
    Ser Pro Arg Trp Leu Val Glu Lys Gly Lys Leu Glu Glu Ala Arg Glu
        210              215              220
40 Val Leu Ala Lys Leu Arg Gly Val Glu Asp Val Asp Gln Glu Ile Gln
        225              230              235              240
    Glu Ile Lys Ala Glu Leu Glu Ala Thr Val Ser Glu Glu Lys Ala Gly
        245              250              255
    Lys Ala Ser Trp Gly Glu Leu Phe Arg Gly Arg Thr Arg Pro Lys Val
        260              265              270
45 Arg Gln Arg Leu Leu Met Gly Val Met Leu Gln Ala Phe Gln Gln Leu
        275              280              285
    Thr Gly Ile Asn Ala Ile Phe Tyr Tyr Ser Pro Thr Ile Phe Lys Ser
        290              295              300
    Val Gly Val Ser Asp Ser Val Ala Ser Leu Leu Val Thr Ile Ile Val
        305              310              315              320
50 Gly Val Val Asn Phe Val Phe Thr Phe Val Ala Leu Ile Phe Leu Val
        325              330              335
    Asp Arg Phe Gly Arg Arg Pro Leu Leu Leu Gly Ala Ala Gly Met
        340              345              350
    Ala Ile Cys Phe Leu Ile Leu Gly Ala Ser Ile Gly Val Ala Leu Leu

```

EP 1 266 906 A1

		355				360				365				
	Leu	Leu	Asn	Lys	Pro	Lys	Asp	Pro	Ser	Ser	Lys	Ala	Ala	Gly Ile Val
		370					375					380		
5	Ala	Ile	Val	Phe	Ile	Leu	Leu	Phe	Ile	Ala	Phe	Phe	Ala	Leu Gly Trp
	385					390					395			400
	Gly	Pro	Ile	Pro	Trp	Val	Ile	Leu	Ser	Glu	Leu	Phe	Pro	Thr Lys Val
					405					410				415
	Arg	Ser	Lys	Ala	Leu	Ala	Leu	Ala	Thr	Ala	Ala	Asn	Trp	Leu Ala Asn
				420					425					430
10	Phe	Ile	Ile	Gly	Phe	Leu	Phe	Pro	Tyr	Ile	Thr	Gly	Ala	Ile Gly Leu
			435					440					445	
	Ala	Leu	Gly	Gly	Tyr	Val	Phe	Leu	Val	Phe	Ala	Gly	Leu	Leu Val Leu
		450					455					460		
	Phe	Ile	Leu	Phe	Val	Phe	Phe	Val	Pro	Glu	Thr	Lys	Gly	Arg Thr
	465				470					475				480
15	Leu	Glu	Glu	Ile	Glu	Glu	Leu	Phe						
					485									

	<210>	6
	<211>	26
20	<212>	PRT
	<213>	Artificial Sequence
	<220>	
	<223>	Consensus sequence for Prosite PS00217
25	<221>	VARIANT
	<222>	1
	<223>	The Leu in the first position can be Ile, Val, Met, or Phe.
30	<221>	VARIANT
	<222>	2
	<223>	The amino acid in the second position can be any amino acid.
35	<221>	VARIANT
	<222>	4
	<223>	The Leu in the fourth position can be Ile, Val, Met, Phe, or Ala.
40	<221>	VARIANT
	<222>	(5)...(6)
	<223>	The amino acids in the fifth and sixth positions can be any amino acid.
45	<221>	VARIANT
	<222>	(8)...(15)
	<223>	The amino acids in the eighth through fifteenth positions can be any amino acid.
50	<221>	VARIANT
	<222>	(16)...(16)
55	<221>	VARIANT
	<222>	(16)...(16)

<223> The Leu in the sixteenth position can be Ile, Phe,
or
Tyr.

<221> VARIANT
<222> (17)...(18)
<223> The amino acids in the seventeenth and eighteenth
positions can be any amino acid.

<221> VARIANT
<222> (19)...(19)
<223> The Glu in the nineteenth position can be Gln.

<221> VARIANT
<222> (20)...(25)
<223> The amino acids in the twentieth through
twenty-fifth positions can be any amino acid.

<221> VARIANT
<222> (26)...(26)
<223> The Arg in the twenty-sixth position can be Lys.

<400> 6
Leu Xaa Gly Leu Xaa Xaa Gly Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Leu
1 5 10 15
Xaa Xaa Glu Xaa Xaa Xaa Xaa Xaa Xaa Arg
20 25

Claims

1. An isolated nucleic acid molecule selected from the group consisting of:

- a) a nucleic acid molecule comprising a nucleotide sequence which is at least 80% identical to the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3;
- b) a nucleic acid molecule comprising a fragment of at least 1000 nucleotides of the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO:3;
- c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2;
- d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO: 2; and
- e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID NO: 1, 3, or a complement thereof, under stringent conditions.

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

- a) a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO:3; and
- b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO:2.

3. The nucleic acid molecule of claim 1 or 2, further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1, 2 or 3, further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of any of claims 1 to 4.

6. The host cell of claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of any of claims 1 to 4.

8. An isolated polypeptide selected from the group consisting of:

a) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 80% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO:3, or a complement thereof;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO: 1, SEQ ID NO:3, or a complement thereof under stringent conditions; and

c) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID NO:2.

10. The polypeptide of claim 8 or 9 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of claim 8, 9 or 10.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID NO:2;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID NO:2, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO:2; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO:2, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID NO:1, SEQ ID NO:3, or a complement thereof under stringent conditions;

comprising culturing the host cell of claim 5, 6 or 7 under conditions in which the nucleic acid molecule is expressed.

13. A method for detecting the presence of a polypeptide of claim 8, 9 or 10 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 8, 9 or 10, as the case may be; and

b) determining whether the compound binds to the polypeptide in the sample.

14. The method of claim 13, wherein the compound which binds to the polypeptide is an antibody.

15. A kit comprising a compound which selectively binds to a polypeptide of claim 8, 9 or 10 and instructions for use.

16. A method for detecting the presence of a nucleic acid molecule of any of claims 1 to 4 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

17. The method of claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

18. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of any of claims 1 to 4 and instructions for use.

19. A method for identifying a compound which binds to a polypeptide of claim 8, 9 or 10 comprising the steps of:

- a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8, 9 or 10 as the case may be with a test compound; and
- b) determining whether the polypeptide binds to the test compound.

5 **20.** The method of claim 19, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

- a) detection of binding by direct detecting of test compound/polypeptide binding;
- b) detection of binding using a competition binding assay;
- 10 c) detection of binding using an assay for 46584 mediated signal transduction.

15 **21.** A method for modulating the activity of a polypeptide of claim 8, 9 or 10 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8, 9 or 10 as the case may be with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

20 **22.** A method for identifying a compound which modulates the activity of a polypeptide of claim 8, 9 or 10, comprising:

- a) contacting a polypeptide of claim 8, 9 or 10 as the case may be with a test compound; and
- 25 b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

30

35

40

45

50

55

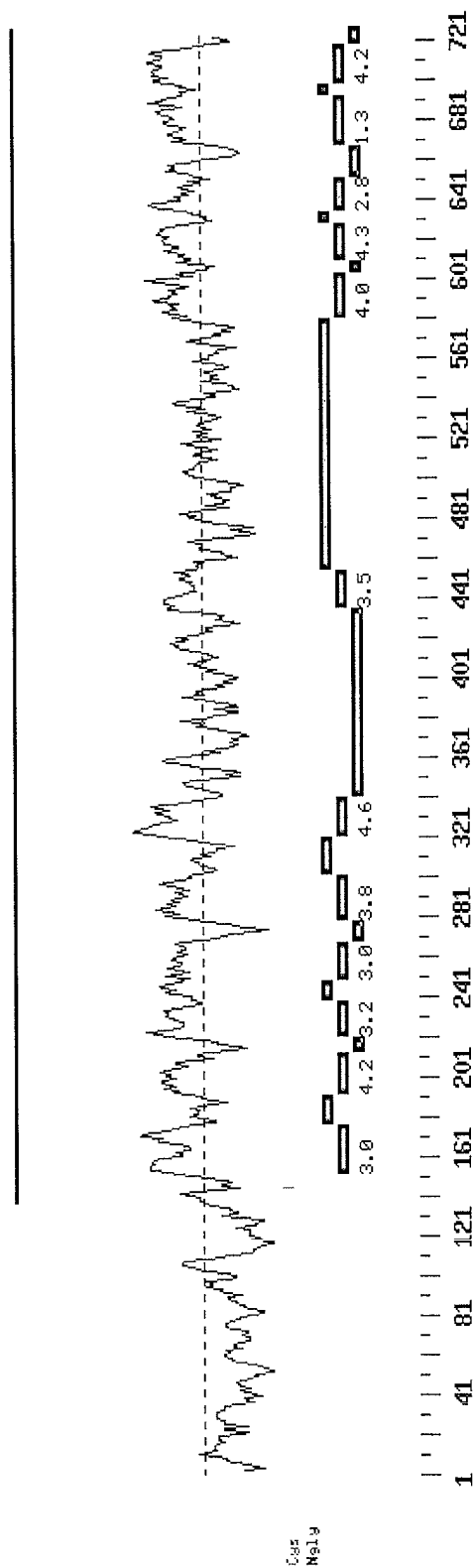


FIGURE 1



European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number

which under Rule 45 of the European Patent Convention shall be considered, for the purposes of subsequent proceedings, as the European search report

EP 02 25 4112

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
D, X	JANZ ROGER ET AL: "SVOP, an evolutionarily conserved synaptic vesicle protein, suggests novel transport functions of synaptic vesicles." JOURNAL OF NEUROSCIENCE, vol. 18, no. 22, 15 November 1998 (1998-11-15), pages 9269-9281, XP002211336 ISSN: 0270-6474 * page 9280, line 1-5; figures 1,4 * -& DATABASE EMBL 'Online! 23 November 1998 (1998-11-23) JANZ R. ET AL, : "Rattus norvegicus synaptic vesicle protein 2C (SV2C) mRNA, complete cds" Database accession no. AF060174 XP002211390 --- -/--	1-22	C07K14/47 C07K16/18 C12N15/12 C12N5/10 C12Q1/68 G01N33/68
			TECHNICAL FIELDS SEARCHED (Int.Cl.7) C07K C12N C12Q G01N
INCOMPLETE SEARCH The Search Division considers that the present application, or one or more of its claims, does/do not comply with the EPC to such an extent that a meaningful search into the state of the art cannot be carried out, or can only be carried out partially, for these claims. Claims searched completely : Claims searched incompletely : Claims not searched : Reason for the limitation of the search: Although claim 21 could be directed to a method of treatment of the human/animal body (Article 52(4) EPC), the search has been carried out and based on the alleged effects of the compound/composition.			
Place of search		Date of completion of the search	Examiner
MUNICH		16 September 2002	Mabit, H
CATEGORY OF CITED DOCUMENTS X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document			

EPO FORM 1503 03.82 (P04097)



European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number
EP 02 25 4112

DOCUMENTS CONSIDERED TO BE RELEVANT			CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	
X	KIKUNO R ET AL: "PREDICTION OF THE CODING SEQUENCES OF UNIDENTIFIED HUMAN GENES. XIV. THE COMPLETE SEQUENCES OF 100 NEW CDNA CLONES FROM BRAIN WHICH CODE FOR LARGE PROTEINS IN VITRO" DNA RESEARCH, UNIVERSAL ACADEMY PRESS, JP, vol. 6, 1999, pages 197-205, XP000852618 ISSN: 1340-2838 * gene product KIAA1054, table 2, page 200* * page 204, column 2, line 8-10; table 1 *	1-22	
Y	BINDRA PAVELJIT S ET AL: "Conservation of the amino acid sequence of SV2, a transmembrane transporter in synaptic vesicles and endocrine cells." GENE (AMSTERDAM), vol. 137, no. 2, 1993, pages 299-302, XP001093607 ISSN: 0378-1119 *Fig 1 legend* * page 302, column 2, paragraph 5; figure 2 *	1-22	TECHNICAL FIELDS SEARCHED (Int.Cl.7)
Y	JANZ R: "SV2 as a calcium dependent regulator of neurotransmission." SOCIETY FOR NEUROSCIENCE ABSTRACTS, vol. 26, no. 1-2, 2000, pages Abstract No.-129.1, XP002211337 30th Annual Meeting of the Society of Neuroscience; New Orleans, LA, USA; November 04-09, 2000 ISSN: 0190-5295 * the whole document *	1-22	
	----- -/--		



European Patent
Office

PARTIAL EUROPEAN SEARCH REPORT

Application Number
EP 02 25 4112

DOCUMENTS CONSIDERED TO BE RELEVANT			CLASSIFICATION OF THE APPLICATION (Int.Cl.7)
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	
P,X	WO 01 75067 A (HYSEQ INC ;LIU CHENGHUA (US); TANG Y TOM (US); DRMANAC RODOJE T (U) 11 October 2001 (2001-10-11) *SEQ ID N° 18152, SEQ ID N° 21738, SEQ ID N° 52106, SEQ ID N° 48520*	1,8	
			TECHNICAL FIELDS SEARCHED (Int.Cl.7)

**ANNEX TO THE EUROPEAN SEARCH REPORT
ON EUROPEAN PATENT APPLICATION NO.**

EP 02 25 4112

This annex lists the patent family members relating to the patent documents cited in the above-mentioned European search report. The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

16-09-2002

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0175067 A	11-10-2001	AU 3486501 A	14-08-2001
		AU 3995501 A	12-09-2001
		AU 4925101 A	15-10-2001
		AU 4972801 A	15-10-2001
		AU 5119401 A	15-10-2001
		AU 5521401 A	15-10-2001
		WO 0157266 A1	09-08-2001
		WO 0164839 A2	07-09-2001
		WO 0175067 A2	11-10-2001
		WO 0175064 A2	11-10-2001
		WO 0174836 A1	11-10-2001
		WO 0175093 A1	11-10-2001
		US 6436703 B1	20-08-2002
		US 2002061567 A1	23-05-2002
